Child L2, Adult L2, Child L1: Differences and Similarities

A study on the acquisition of direct object scrambling in Dutch

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6000
e-mail: lot@let.uu.nl
http://wwwlot.let.uu.nl/

Cover illustration: *by Alison Unsworth*

ISBN 90-76864-87-X

NUR 632

Child L2, Adult L2, Child L1: Differences and Similarities

A study on the acquisition of direct object scrambling in Dutch

Vroege Tweede Taalverwerving, Late Tweede Taalverwerving,
Vroege Eerste Taalverwerving: Verschillen en Overeenkomsten
Een onderzoek naar de verwerving van
scrambling van het lijdend voorwerp in het Nederlands
(met een samenvatting in het Nederlands)

PROEFSCHRIFT
ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof. Dr. W.H.Gispen,
ingevolge het besluit van het Collge voor Promoties
in het openbaar te verdedigen
op woensdag 21 december 2005
des middags te 14:30 uur

door

Sharon Unsworth

geboren op 10 oktober 1974 te Preston, Verenigd Koninkrijk

Promotores: Prof. dr. P.C.A. Coopmans
Faculteit der Letteren, Universiteit Utrecht

Prof. dr. B.D. Schwartz
College of Languages, Linguistics and Literature,
University of Hawai'i at Mānoa

ACKNOWLEDGEMENTS

It was probably my primary school teacher, Mrs. Sykes, who first sparked my interest in language. Every week she would pull out items from her suitcase of props, asking *Qu'-est-ce que c'est?* She happened to speak French and luckily, we all got to profit from this. Although the serious interest in linguistics and language acquisition came later, at university, my own early exposure to a second language is clearly reflected in my choice of topic for this thesis.

The research which has contributed to the book you see before you started life at the University of Durham and I would like to start by thanking Mike Davenport and the Department of English Language and Linguistics (sadly soon to be no more) for their support (financial and otherwise) during the initial stages of my project and in my transition to The Netherlands. Thanks also to Bonnie D. Schwartz and Joe Emonds for suggesting that I apply for a Marie Curie fellowship at Utrecht institute of Linguistics – something which has turned out to have a significant impact on my life, and not just academically – and to UiL OTS for taking me on as an *AiO*. It was also Bonnie who suggested that I work on scrambling. (Until then, I thought that this was something you only did to eggs.)

Fortunately, transferring from Durham to Utrecht did not put an end to Bonnie's role as supervisor. Bonnie's ideas and insights have had a great influence on my work and they constitute the starting point for this thesis. I would like to thank her for teaching me to conduct thorough, precise research, for her painstaking attention to detail, her enthusiasm and for her confidence in me. I would also like to thank my local supervisor, Peter Coopmans, for believing in me, for helping me to step back and see the big picture, and for listening to my concerns and taking them seriously.

I am indebted to Aafke Hulk for regularly making time to listen to my ideas, for her enthusiasm and encouragement, and for somehow having this uncanny knack of always asking the kind of questions which spark some (usually very useful) connection somewhere deep down in my brain. Thanks also to Elma Blom for *het meedenken* about the problems and pitfalls of child L2 ~ adult L2 comparisons, and to the ever-patient and always smiling Hugo Quené for help with all things statistical, especially with the principal components analysis which I used for the proficiency measure in Chapter 4. A special word of thanks also goes to Cécile De Cat for her support and enthusiasm, for her excellent company during numerous conferences and workshops (where will we go next?) and, as part of a very enjoyable collaboration, for teaching me to be careful in expressing my ideas. There are two names which pop up time and time again in this thesis, namely those of Irene Krämer and Jeanette Schaeffer. I would like to thank them both for the numerous fruitful discussions we've had, for their enthusiasm and interest, and for generously making test materials and individual data available to me.

I have also benefited from fruitful discussions with and/or useful advice and feedback from the following people: Sonja van Boxtel, Harald Clahsen, Norbert Corver, Kamil Ud Deen, Laurent Dekydtspotter, Helen de Hoop, Holger Hopp, Anna Kijak, Karen Miller, Rick Nouwen, Bill Philip, Eddy Ruys, Roumyana Slabakova, Antonella Sorace, Rex Sprouse, Henriëtte de Swart, Kriszta Szendrői, Yi-ching Su, Fred Weerman and Melinda Whong-Barr. I would also like to thank Yi-ching Su for kindly making her individual data available to me.

Parts of this thesis have been presented at various conferences and workshops over the past few years. I am very grateful to UiL-OTS for making this financially possible. A longer trip to Boston and Honoulu, USA, in 2004 was co-financed by NWO (Netherlands Organisation for Scientific Research), which I gratefully acknowledge. I have particularly benefited from comments and feedback during presentations for the Edinburgh Developmental Linguistics group and the University of Essex and the University of Hawai'i Linguistics Colloquia.

Several people have read and provided feedback on earlier versions of one or more chapters of this thesis, for which I am grateful: Elma Blom, Cécile De Cat, Berit Gehrke, Rick Nouwen and Eddy Ruys.

Thanks to various colleagues and/or friends, in various locations, for making life both inside and outside of linguistics such a pleasant place to be: Berit (the perfect officemate!), Brigit, John, Willemijn, Gert Jan, Henri, (BH) Simon, Kai, Verena, Helen, Dom, (LH) Simon, Barbara, Akira, Melinda, Lieve, Stéph, Ellie, Helen, Patrick H, Elise, Shakuntala, Øystein, Patrick B, Rick, Olga, Silke. And I should also not forget allotment 31-A at the *Groene Lunet* and the Boretti V-946, which both played a crucial role in keeping me sane whilst writing up.

This thesis would not exist at all if it were not for the willingness of the many test subjects to take part in my experiments. Thanks to all the participants and to the following teachers and schools for their cooperation: Mariet Mennen and the Regional International School of Eindhoven; Lorraine Boyle and De Blijberg, International Primary School, Rotterdam; Patricia Murphy and the International School of Amsterdam; Jane Forrest and Rotterdam International Secondary School; Cheryl Embleton and Hubert Keulers and Elckerlyc Leiden International Primary School; Chris Ellison and Het Rijnlands Lyceum

International School, Oegstgeest; Marga Akerboom and the American International School of Rotterdam; Rick van Vliet and the International School of The Hague; Sannerijn Jansen and the International School of Hilversum; Basisschool De Rank, Koudekerk aan de Rijn; Basisschool De Boschuil, Eindhoven; Christiaan Huygens College, Eindhoven; and St. Vituscollege, Naarden. I have also been lucky enough to have the assistance of various students and *stagiaires*: Renate van den Berg, Neely Anne Davidse, Wil van Goch, Hiske van der Meulen and Arjen Zondervan. A special thanks goes to Wil for her enthusiasm, dedication and *gezelligheid*.

Thanks to my family in Preston, Newcastle, Amsterdam, Heerlen and Zeist for their love and support.

Thanks especially to Alison and Laurence for their fine company (and facilities) during many a top-tastic holiday(ette) at the RRHs – always a welcome break from linguistics. An extra special thank-you to Alison for designing the front cover of this book, for teaching me how to draw frogs over the phone, for her last minute help taming Microsoft Word and most importantly, for always looking after me.

Were it not for Rick, I can safely say that I would now be reduced to an emaciated wreck, collapsed in front of my computer, surrounded by half-empty cups of tea in various stages of decay. I want to thank him for keeping me fed, watered, happy and just generally very well looked-after (even from a distance) during these past few months. I would also like to thank him for listening to me and asking the right questions, for his tutorials in semantics, for making sure (when necessary) that I 'get a grip', and most of all, for his unwavering belief in me.

Knowing that there are people who believe in me has been a real help while writing this thesis. It is in appreciation of this fact, and with great sadness, that I dedicate this book to the memory of my dear friend, Ellie Price.

TABLE OF CONTENTS

INTRODUCTION

This thesis compares three different groups of language learners – non-native (L2) children, L2 adults and first language (L1) children – in their acquisition of the same target language property, namely direct object scrambling in Dutch. These groups are compared and contrasted in their linguistic development to answer the question of whether, broadly speaking, child L2, adult L2 and child L1 acquisition essentially involve the same process(es).

The question of whether adult L2 acquisition is constrained by the same innate language mechanism as L1 acquisition, i.e. Universal Grammar (UG), has been the focus of much research during the past twenty years in the field of (generative) L2 acquisition. Many of the studies addressing this question directly compare adult L2 acquisition with child L1 acquisition. Taking centre stage in the cross-learner comparison in this thesis, however, is child L2 acquisition. On the one hand, L2 children are compared with L2 adults, and on the other, they are compared with L1 children. In Chapter 1, I argue that not only is child L2 acquisition of interest in and of itself, but it can also be informative with respect to the acquisition process in the other two groups, L2 adults and L1 children. It is this focus on child L2 acquisition which distinguishes this thesis from earlier work in the field. The present study is one of the first to directly compare L2 children, L2 adults and L1 children using exactly the same experimental methodology with each of these learner groups. In addition to presenting new data from these three groups, the thesis also highlights some of the conceptual and methodological issues which need to be taken into consideration in any such cross-group comparative study.

The importance of child L2 acquisition stems from crucial similarities and differences it shares with each of these comparator groups. Two factors are important in this regard, namely age and L1 transfer. L2 children and L2 adults are by definition similar in that both groups have knowledge of another language and this allows for the potential of L1 transfer. These two groups differ with respect to age, however, because the L2 children, if tested when still children, are younger than L2 adults and in this sense, they are (to a certain extent, at least) more similar to L1 children. In this thesis, following Schwartz (1992), the possible effects of these similarities and differences are explored in the domain of developmental sequences. Specifically, this thesis asks whether L2 children, L2 adults and L1 children pass through the same developmental sequences in their acquisition of direct object scrambling in Dutch.

In Dutch, direct object NPs may appear to the left or to the right of sentential adverbs, such as *twee keer* 'twice', and negation. This is illustrated for indefinite objects in (1) and (2).

(1) a.

Het meisje	heeft	twee keer	een aap	gekieteld	[non-scrambled]
the girl	has	two times	a monkey	tickled	

'The girl tickled a(ny) monkey twice.'

b.

Het meisje	heeft	een aap	twee keer	gekieteld	[scrambled]
the girl	has	a monkey	two times	tickled	

'The girl tickled a (certain) monkey twice.'

(2) a.

De jongen	heeft	geen (niet + een)	vis	gevangen	[non-scrambled]
the boy	has	no (not + a)	fish	caught	

'The boy didn't catch a(ny) fish.'

b.

De jongen	heeft	een vis	niet	gevangen	[scrambled]
the boy	has	a fish	not	caught	

'The boy didn't catch a (certain) fish.'

In (1) and (2), the object moves from its base position directly adjacent to the verb in the a-sentences to the 'scrambled' position to the left of the adverb/negation in the b-sentences. The position of the indefinite object relative to the adverb or negation has an effect on how it is interpreted. When it appears to the right of the adverb or negation in non-scrambled position, the indefinite object receives a non-specific interpretation, whereas when it occurs to the left of the object or negation in scrambled position, it has a specific interpretation. The scrambled sentence in (1)-b therefore singles out a certain monkey which the girl tickled on two different occasions, whereas the indefinite object in the non-scrambled sentence in (1)-a can refer to any monkey. Likewise, the indefinite object in the scrambled sentence in (2)-b is interpreted as referring to a certain fish which the boy did not catch, whereas the indefinite object in the non-scrambled sentence in (2)-a can refer to any fish, and given that it is in the scope of negation, the sentence means that the boy did not catch any fish (at all). Scrambling of indefinite object NPs is thus viewed as a way of disambiguating meaning. In Chapter 2, scrambled indefinite objects are analysed as being of a different semantic type from non-scrambled indefinite objects (de Hoop 1992; Van Geenhoven 1998).

The acquisition of scrambling of definite object NPs across negation is also addressed. Non-scrambled definite objects in negative sentences are usually interpreted as expressing contrastive negation and scrambled definite objects generally receive a sentential negation interpretation. This is illustrated in (3)-a and (3)-b, respectively. Both sentences are given with neutral stress; main sentence stress (and *not* contrastive stress) is indicated by SMALL CAPITALS.

(3) a.

Het meisje	heeft	niet	het PLAATJE	nagetekend
the girl	has	not	the picture	copied

'The girl hasn't copied the PICTURE.'

b.	Het meisje	heeft	het plaatje	niet	NAgetekend
	the girl	has	the picture	not	copied
	'The girl hasn't COPIED the picture.'				

The non-scrambled sentence in (3)-a is interpreted as expressing contrastive negation on the object (or the VP), that is, it implies that there is something other than the picture which the girl did copy. This contrastive interpretation is not possible for the scrambled sentence in (3)-b; rather, this sentence is interpreted as expressing sentential negation. This (non-truth-conditional) difference between scrambled and non-scrambled definite objects in negative sentences is analysed as the result of the interaction between negation and focus (following Neeleman and Reinhart 1998; Reinhart 1995).

This thesis contributes to the increasing body of research on L2 acquisition at the syntax-semantics interface by investigating whether L2 children and adults, whose L1, English, does not have scrambling, can come to know the interpretive constraints associated with the word order alternations in (1) through (3). In particular, on the basis of cross-sectional, quantitative, experimental data, it is determined whether these two groups of L2ers come to know that (i) when a sentential rather than a contrastive negation interpretation is intended, definite object NPs are generally scrambled, and (ii) the non-specific interpretation of the indefinite object is ruled out when it occurs in scrambled position. In Chapter 3, I argue that the acquisition of this latter property of Dutch constitutes a poverty-of-the-stimulus problem for English-speaking (child and adult) L2ers: the fact that the scrambled indefinite objects in (1)-b and (2)-b cannot be associated with a non-specific interpretation cannot be induced from the L2 input, it cannot be transferred from the L1, and it is unavailable in the L2 classroom. It is therefore concluded that if the L2ers in this study demonstrate targetlike knowledge of these constraints, their interlanguage grammars must be constrained by in the same way as L1 acquisition is.

To investigate whether this is the case, two sets of experiments were carried out, the results of which are presented in Chapters 5 and 6. These two chapters constitute the empirical heart of the thesis. Chapter 5 reports on a production experiment whose purpose, following Schaeffer (2000), is to establish whether learners produce the targetlike form (scrambled vs. non-scrambled indefinite) in the relevant context (specific vs. non-specific). This experiment also determines whether learners produce scrambled definite objects in negative sentences when a sentential negation interpretation is intended. Chapter 6 presents a series of three comprehension experiments which, following Krämer (2000), investigates whether learners assign the targetlike interpretation (specific vs. non-specific) to the relevant form (scrambled vs. non-scrambled). All of the experiments are carried out with three learner groups, that is, L2 children, L2 adults and L1 children, as well as with native adult controls. In order to determine the developmental sequences for the L2ers, subjects were divided into three proficiency groups using proficiency scores based on semi-spontaneous data collected using a picture description task, as outlined in Chapter 4.

Three major findings emerge from Chapters 5 and 6. First, in their production of scrambled definite and indefinite objects, L2 children and L2 adults pass through the same developmental sequence; this sequence only differs from that of the L1 children in terms of the initial stage, which is analysed as the result of L1 transfer. Second, in their comprehension of scrambled objects, L2 children pattern more like L1 children than like L2 adults. More specifically, both groups of child learners appear to be affected by their developing abilities in discourse/pragmatics, which is thought to restrict their scrambling behaviour. Third, English-speaking L2 children and L2 adults are able to overcome the poverty-of-the-stimulus to acquire the interpretive constraints on scrambled indefinite objects in Dutch.

These findings furthermore suggest that for L1 and L2 children, there is a discrepancy between production and comprehension of scrambled indefinites, with targetlike production in place before targetlike comprehension. In Chapter 7, this pattern is confirmed in the data from a subset of L2 children who took part in both the production and the comprehension experiments: while there are several subjects who have targetlike production and non-targetlike comprehension, there are no subjects with the reverse pattern, that is, there are no subjects who have targetlike comprehension but non-targetlike production. Various explanations are explored for this observation. These include methodological factors, such as the set-up of the comprehension experiments, and the discourse/pragmatic factors mentioned above, such as limited discourse integration (Krämer 2000).

The observation that (i) L2 children and adults are able to come to know the interpretive constraints on scrambling and (ii) that in production at least, these two groups pass through the same developmental sequence, suggests that, on the assumptions which are laid out in Chapter 1, (adult) L2 acquisition is constrained by UG. In the final part of the thesis, I explore how the interpretive constraints on scrambling might be acquired. I advocate the view that learners make use of a Blocking Principle (Williams 1997) which ensures that once the scrambled and non-scrambled forms are related to each other by means of syntactic movement, the non-specific reading for scrambled indefinites is ruled out for economy reasons. As the non-scrambled form expresses the non-specific interpretation, it is the preferred expression of this meaning, because given that it does not involve movement, it is 'cheaper'. The application of this principle of language (acquisition) is argued to guide the learner to targetlike knowledge of scrambling, although other (non-domain-specific) factors may still lead to non-targetlike responses under certain (experimental) circumstances.

CHAPTER 1

CHILD SECOND LANGUAGE ACQUISITION

Introduction

Child L2 acquisition constitutes the heart of this thesis. It is compared with, on the one hand, adult L2 acquisition and, on the other, child L1 acquisition. These three groups share various similarities and differences, the conceptual and empirical basis for which is presented in this chapter. This chapter furthermore provides the motivation for the comparative approach adopted here. It also presents a detailed review of recent research into child L2 acquisition. One of the clearest results of this review is the finding that like adult L2 acquisition, child L2 acquisition is characterised by L1 transfer, which sets both these groups apart from L1 children. In several other areas, however, the available data are too few to draw any firm conclusions, which further serves to highlight the need for studies such as the one carried out here. In the course of this critical appraisal of the literature, several important methodological issues are encountered and these will serve to motivate the design of the present study.

The chapter is organised as follows. First, the concept of child L2 acquisition is defined in §1.1. Subsequently, the motivation for research into child L2 acquisition and what this can tell us about adult L2 and child L1 acquisition is presented in §1.2. Section 1.3 provides a brief introduction to the research which has been carried out on this topic to-date and this research is then explicated and reviewed in the subsequent two sections: in §1.4, research which compares child L2 acquisition with adult L2 acquisition is presented, and in §1.5, the focus falls on child L2 acquisition compared with child L1 acquisition. The chapter draws to a close with a list of desiderata for cross-group comparisons such as the one which will be executed here.

1.1 Defining child L2 acquisition

In this thesis, the term *child L2 acquisition* is used to distinguish successive child bilingualism from simultaneous child bilingualism, on the one hand, and successive child bilingualism from successive adult bilingualism, on the other. Child L2 acquisition (successive child bilingualism) is considered to differ from bilingual L1 acquisition (2L1A; simultaneous child bilingualism), which is defined as the exposure to two languages from birth or soon afterwards (see Meisel 2004 for a recent overview of the 2L1A literature). Child L2 acquisition will thus be used to denote L2ers whose first exposure to the second language occurs at a point at which the bulk of their first (native) language is already in place.[1] With very few empirical studies addressing the issue of where the line should be drawn between these two types of acquisition (but see Kroffke and Rothweiler to appear), whichever age is chosen will, to a certain extent, remain rather arbitrary (McLaughlin 1978). McLaughlin (1978) suggests a cut-off age of three years, on the grounds that an L2 child's first language is relatively well established by this point. However, as Lakshmanan (1995:322 , fn.5) indicates (even though she does adopt McLaughlin's criterion herself), this is probably too early, given that not all complex properties of language have been acquired by this age. In the present study, I will assume a cut-off point of four years. At this age, we can assume most (purely) grammatical principles (and, for example, the phonology) of the first language to be in place (e.g. Goodluck 1986; Guasti 2002). Ultimately, of course, this cut-off point should be determined empirically.

The present study compares child L2 acquisition with adult L2 acquisition. To do this, it is necessary to define where the difference lies between these two groups. In other words, at what age does child L2 acquisition stop being child L2 acquisition and become an adult L2 acquisition? The answer to this question is intrinsically linked to the debate about whether there is a critical period in language acquisition. This debate, which will be discussed in more detail in §1.2.1.1, is far from resolved and various ages have been put forward as the start of a decline in the ability to reach nativelike levels of L2 proficiency:[2] five years (Krashen 1973), seven years (DeKeyser 2000; Johnson and Newport 1989), eight years (Bialystok and Miller 1999), nine years (Penfield and Roberts 1959), puberty (Lenneberg 1967) and 15 years (Long 1990). Following Schwartz (2004), I will adopt the conservative value of eight years. This is because, as Schwartz (2004) notes, in both Johnson and Newport (1989; 1991) and DeKeyser (2000), children who started to acquire their L2 before this age were observed to fall within the range of native-speakers on various tasks testing a variety of morphosyntactic phenomena.[3] The L2 child is therefore defined as in (1).

[1] In other words, L2ers are classified according to the point at which the L2 acquisition process begins. No claims are made about the nature of their end-state grammars. Hence, irrespective of whether they could be described as (functionally) bilingual or nativelike, this type of language learner will still be referred to as an L2er.

[2] Note that in and of itself, a decline in nativelike attainment across an increasing age span does not necessarily constitute evidence for a critical period for language acquisition. In order for this to be the case, such a decline should (at the very least) exhibit a discontinuous rather than continuous function (Bialystok and Miller 1999).

[3] Note that although Johnson and Newport's work (1989; 1991) has been criticised on various grounds (see below), this criticism has not specifically challenged the claim that in their data, age seven differentiates nativelike L2ers from non-nativelike L2ers. Note furthermore that using these data to define child L2 acquisition does not require a commitment to the conclusion these authors draw, that is, that there is a critical period for language acquisition.

(1) Definition of L2 child
A non-native acquirer whose initial exposure to the target language is between the ages of four and seven years.

Consequently, the L2 adult is defined as follows:

(2) Definition of L2 adult
A non-native acquirer whose initial exposure to the target language is at an age of eight years or older.

It might be objected that the group of adult L2ers resulting from the definition in (2) will be rather diverse, in the sense that it will potentially contain learners whose ages (at first exposure) range from eight to 80. In order to ensure that the conclusions drawn in the present study do not crucially depend on the specific definitions of the L2 child and L2 adult given above, additional analyses of the experimental data will be conducted where, for example, the lower age limit for adult L2 acquisition is raised to exclude pre-pubertal subjects.

Note, finally, that although child L2 acquisition is defined here in terms of age at first exposure, age at time of testing will also be an important factor in this thesis, especially in setting up the predictions made for the comprehension experiment in Chapter 6 and in devising the proficiency score in Chapter 4.

The following section presents the motivation for studying child L2 acquisition and its relevance to the fields of L1 and L2 acquisition research.

1.2 The importance of child L2 acquisition

Child L2 acquisition has been argued to be informative with respect to the nature of both child L1 acquisition and adult L2 acquisition (Lakshmanan 1995; Schwartz 1992; 2003; 2004). By adopting the comparative method proposed by the latter author, that is by comparing these three different groups of language learners – L2 children, L2 adults and L1 children – in their acquisition of a particular property of the target language, this thesis asks whether child L2, adult L2 and child L1 acquisition involve the same process(es). The rationale behind this comparison and how (child/adult) L2 acquisition and L1 acquisition may or may not involve the same process is spelled out in the remainder of this chapter and the specific predictions to be tested are formulated in Chapter 3.

The following section considers how child L2 acquisition can be informative with respect to adult L2 acquisition, outlining the logic of the child L2 ~ adult L2 comparison. Studies which make such a comparison will be reviewed in §1.4.

1.2.1 What child L2 can tell us about adult L2 …

The first way in which the study of child L2 acquisition may be beneficial to research on adult L2 acquisition is quite simply by providing new data with which to test theories of L2 acquisition based on adult data only. For example, several researchers (Gavruseva and Lardiere 1996; Grondin and White 1996; Haznedar 2001; 2003; Jin 2003) have tested the predictions of the approach to L2 acquisition known as Minimal Trees (Vainikka and Young-Scholten 1994; 1996a) using child L2 data.

Child L2 acquisition may also be informative with respect to the nature of the adult L2 acquisition process. Comparisons of L2 children and L2 adults usually focus on whether L2 acquisition proceeds in the same fashion for both groups, that is, whether, as a result of biological and/or cognitive and/or sociological factors, the acquisition of a second language as an adult is fundamentally different (Bley-Vroman 1989) from the acquisition of a second language as a child.

Before outlining the rationale for the child L2 ~ adult L2 comparison in the present study in §1.2.1.2, I briefly review some of the major findings of research on this topic and the issues involved.[4]

1.2.1.1 Maturational constraints in L2 acquisition

Although the research questions addressed in this thesis will not be explicitly phrased in such a way, the issue of maturational constraints in L2 acquisition is often phrased in terms of whether or not there is a critical period for language acquisition.[5] In its most theory-neutral form, the Critical Period Hypothesis (CPH) states that 'there is a limited developmental period during which it is possible to acquire a language, be it L1 or L2, to normal, nativelike levels' (Birdsong 1999:1).[6]

In their comprehensive review of maturational constraints in L2 acquisition, Hyltenstam and Abrahamsson (2003) consolidate the empirical observations made in the CPH literature into three broad positions:

---

[4] For a more complete overview of the extensive literature on maturational constraints in L2 acquisition, see Birdsong (1992), Harley (1997), Hyltenstam and Abrahamsson (2003), Long (1990), Scovel (2000) and Singleton (2004).

[5] The issue of whether there is a critical period for language acquisition is clearly implied in the central research question of this thesis. My research is not explicitly phrased in these terms, however, because it does not address the question of whether all the geometric features of a critical period are present in the data which are collected and whether, if they are, these have a biological foundation (see e.g.Van Boxtel 2005 for recent discussion on this topic).

[6] As Birdsong goes on to note, although there is no single CPH, each of its various formulations predicts a non-nativelike end-state for older/adult L2ers and seeks to explain their widespread failure to attain nativelike proficiency in terms of developmental factors which affect the human species as a whole. These factors may be biological (e.g. Lenneberg 1967; Pulvermüller and Schumann 1994), cognitive (e.g. (Felix 1985; Newport 1990) or affective (Krashen 1982; Schumann 1975).

(3) A. nativelike L2 proficiency is observed in early/child L2ers only
B. nativelike L2 proficiency is observed in early/child L2ers, and also in individual late/adult L2ers
C. nativelike L2 proficiency is found neither in early/child L2ers nor late/adult L2ers.

Position A, which asserts maturational constraints in L2 acquisition, is best illustrated by perhaps the most cited study testing for a critical period in L2 acquisition: Johnson and Newport (1989). Johnson and Newport (1989) tested native Chinese and Korean speakers (n=46) on a variety of aspects of English grammar (past tense, plural, third person singular, present progressive, determiners, pronominalisation, particle movement, subcategorisation, auxiliaries, yes/no questions, I-questions and word order) using an aural grammaticality judgement task. The subjects had been exposed to English for at least five years and had been resident in the USA for at least three consecutive years. For subjects first exposed to English before puberty ($\leq$ 15 years), there was a significant negative correlation between age of first exposure and task score ($r = -.87$, $p < .01$), whereas amongst post-pubertal acquirers, no such relationship was observed ($r = -.16$, $p > .05$). Johnson and Newport – and since then, many others – interpret these results as evidence for the so-called *the younger, the better* position regarding age and ultimate attainment (see Oyama 1978 for similar results; Patowski 1980; 1990). Although Johnson and Newport's investigation has since been criticised (Bialystok and Hakuta 1994; Bialystok 1997; Kellerman 1995) and in replications and reanalyses their claims have been both confirmed (e.g. DeKeyser 2000; Johnson and Newport 1991) and disconfirmed (e.g. Bialystok and Miller 1999; Birdsong and Molis 2001), most researchers would still concur that early/child L2ers have a long-term advantage over late/adult L2ers.[7]

The set-up of Johnson and Newport's study takes on the typical format of a CPH study: a group of L2ers, whose age of first exposure ranges from child to adult, are tested on a series of properties of the target language to determine whether or not they fall within native-speaker range. Subsequently, correlation co-efficients are calculated to ascertain whether there is a(n inverse) relationship between age at first exposure and ultimate attainment. A second type of CPH study involves specifically selecting late/adult L2ers who are judged to be very advanced (and thus, potentially nativelike) and who have been resident in the host country for a considerable period of time (and can therefore be considered to have reached their end-state). The rationale behind

[7] Several studies have claimed that older L2ers (older children vs. younger children, adults vs. children) have the advantage when it comes to the (initial) rate of acquisition (e.g. Asher and Price 1967; Snow and Hoefnagel-Höhle 1982) (cf. Slavoff and Johnson 1995). As Krashen, Long and Scarcella (1979) point out, however, it is necessary to make a distinction between rate and ultimate attainment. Hyltenstam and Abrahamsson (2003) note that it is not clear what such rate studies can tell us about the critical period. Indeed, the fact that the L2 grammars of certain learners develop more quickly than others does not say anything about the epistemological status of those grammars.

such studies is that any end-state late/adult L2er who achieves a level of proficiency which is within the range of native-speakers constitutes counter-evidence to the CPH. Several studies have found such late/adult L2ers, including, for example, Birdsong (1992), Bongaerts (1999), Ioup, Boustagui, El Tigi and Moselle (1994),Van Boxtel (2005), White and Genesee (1996); cf. Coppetiers (1987). These studies characterise position B in Hyltenstam and Abrahamsson's (2003) classification.[8]

Such a teleological approach to the nature of L2 acquisition has been criticised, however. The logic of such an approach is that if late/adult L2ers are found to reach nativelike levels of proficiency, the critical period (in its strongest sense) cannot hold. Schwartz (1990) points out, however, that in cases where such nativelike proficiency is *not* found, i.e. when native-speaker grammars and late/adult L2 end-state grammars are found to differ, this does not necessarily entail epistemological non-equivalence between the two. It is possible that late/adult L2ers may not have nativelike L2 grammars, but these grammars are nevertheless still constrained in the same way as native-speaker grammars. Schwartz draws an analogy with L1 children who, in situations of language change, do not necessarily exactly reconstruct the grammar of the input providers, and to the grammar of English, which has changed over time (to wit, Old English, Middle English and Modern English). Both cases involve different grammars, but no-one would claim that these grammars are epistemologically non-equivalent.

Returning to Hyltenstam and Abrahamsson's (2003) classification, the third position they describe is the one which they themselves adhere to, namely that nativelike proficiency is found neither in adult/late L2ers nor in early/child L2ers, that is, there is a steady decline in nativelike attainment which starts almost immediately (more or less from birth). This idea is based on three observations: (i) there are early/child L2ers who do not reach nativelike ultimate attainment, (ii) those late/adult L2ers who have been found to reach nativelike levels of proficiency are not nativelike in all domains, and (iii) there is a steady decline in ultimate attainment with increasing age of first exposure, rather than the abrupt cut-off point expected for a critical period (see Hyltenstam and Abrahamsson 2003 for the details of the studies on which these observations are based). There are, however, very few studies which have addressed the issue of whether early/child L2ers always reach nativelike levels of ultimate attainment, and until more data are available, it is difficult to know what to make of this position. Furthermore, and perhaps more importantly, the argument given in the preceding paragraph regarding late/adult L2ers holds for early/child L2ers, too: differences between native-speaker grammars and end-state L2 grammars do not necessarily entail epistemological non-equivalence.

[8] See Sorace (2003) for an overview of issues relating to near-nativeness in adult L2 grammars.

Finally in this section, it should be noted that most studies dealing with age effects in L2 acquisition test early/child L2ers when they are adults as opposed to when they are still young. One of the risks of studies set up in this way is that the adult L2ers, who inevitably have fewer years' of exposure, have not yet reached their end-state. In a (partial) replication of Johnson and Newport (1989), McDonald (2000) found that her L1 Spanish/L2 English adult subjects (n=14, age at time of testing = 21-25 years) were significantly different from both the L2 child subjects (n=14, age at time of testing = 18-26 years) and the native-speaker controls (see Appendix A for more details), but the latter two groups did not significantly differ from each other. There were however significant differences between the two L2 groups in terms of length of exposure: on average, the child L2ers had had 14.8 years' of exposure (range = 4-24 years), whereas the average length of exposure for the adult L2ers totalled just 4.1 years (range: 3-7 years). It is perfectly possible, therefore, as McDonald herself acknowledges, that the differences between these two groups results from differences in length of exposure rather than differences in age at first exposure. Although statistical methods can be employed to factor out this variable (which is what McDonald does, the result being that age at first exposure remains the best predictor of targetlike responses overall), a more clean-cut study would ensure the adult L2ers had had sufficient exposure to guarantee that they had reached the end-state. In such a study, the adult L2 group would therefore have to be significantly older (at time of testing) than the child L2 group. Testing early/child L2ers while they are younger could also potentially alleviate this problem. Given that this thesis will focus on development, it will in any case be crucial to test child L2ers while they are still children.

1.2.1.2 Schwartz (1992): The rationale for the present study

This section outlines the rationale for the present study, based on Schwartz (1992). Schwartz (1992) argues that a comparison of adult L2 with child L2 can inform the debate on the nature of adult L2 acquisition. Specifically, she claims that by comparing developmental sequence data from these two groups, it is possible to decide between a UG-based approach to adult L2 acquisition and a general problem-solving approach. Such a comparison, that is, between the developmental sequences of L2 children and L2 adults, forms the core of this thesis.

Schwartz argues that the two ways in which developmental sequence data are traditionally used in (adult) L2 acquisition research cannot answer the question of whether UG is involved for these learners. The first of these approaches compares the developmental paths of adult L2ers of language X with L1 acquirers of that language. The logic behind this comparison is that if the developmental sequences for these two groups of learners are the same, they must involve the same underlying process, that is, they are constrained by

UG, whereas if the developmental paths are different, the underlying process is assumed to be different. The problem with this method is that any pattern observed in the L2 sequence which may differ from the L1 sequence could potentially be due to L1 influence. The second approach compares developmental sequences of adult L2ers with different L1s; similar patterns are taken to indicate UG-constrained development, whereas different patterns are assumed to confirm the opposite position. Similar developmental sequences do not necessarily have to entail that the same underlying process involves UG, however. Adult L2ers might be using other abilities which they have in common, such as general learning mechanisms. On the other hand, different developmental sequences may also be the result of L1 influence: if the L2ers all have different L1s, and potentially different starting points, it is not necessarily surprising that their developmental paths differ.

Child L2 acquisition, Schwartz claims, may provide the missing link. Her argument works as follows: assuming that (i) child L2 acquisition is driven by UG (based on evidence that young L2ers are generally more successful than adult L2ers, in terms of ultimate attainment (see discussion above)) and (ii) because children (L1 and L2) make use of UG (domain-specific principles), the general learning principles in question are more relevant to adult L2ers, then (iii) comparing developmental sequences of child L2ers with those of adult L2ers, while holding the L1 constant, will provide evidence for or against UG involvement in L2 acquisition.[9] In her (1992) paper, Schwartz argues that if there is (not) L1 transfer, it should (not) be found for both groups. However, in more recent work, given the accumulating evidence that similarly to adult L2 acquisition, child L2 acquisition *is* affected by L1 transfer (see §1.4.2 for discussion of relevant literature), she takes transfer in both groups as given. Similar child L2 ~ adult L2 developmental sequences are taken as evidence for UG involvement in adult L2 acquisition, whereas dissimilar sequences would indicate that adult L2 acquisition stems from something other than UG, i.e. general cognitive mechanisms, and child L2 acquisition stems from UG.[10] This is schematised in (4).

---

[9] The second assumption implies that general learning principles cannot (exclusively) be responsible for L1 acquisition. Whilst in the context of L1 acquisition only, this implication might meet with objections from some quarters, in the present context it seems clear that if general learning principles are responsible for L2 acquisition, they are unlikely to be the same ones (or if they are, they are not used in the same way or to the same extent) as could be claimed to drive L1 acquisition. Otherwise, we would expect L1 and L2 acquisition to pattern much more uniformly, both in terms of development and ultimate attainment.

[10] It might be objected that similar developmental sequences could also be the result of both groups using problem-solving or general learning mechanisms. If this were the case, we would expect the more cognitively mature adult L2ers to be more successful than the less cognitively mature child L2ers (Schwartz 1992:8, fn.6). As outlined above (§1.2.1.1), the L2 data indicate otherwise, however.

(4) a. Child L2 development = Adult L2 development:[a,b]
→ Adult L2 acquisition is constrained by UG
[a] *holding the L1 constant*
[b] *for UG-governed language properties*

b. Child L2 development ≠ Adult L2 development:[a,b]
→ Adult L2 acquisition is not constrained by UG
[a] *holding the L1 constant*
[b] *for UG-governed language properties*

(Schwartz 2004:39)

Note that, as indicated in (4), this only holds for properties of language which are governed by UG. Schwartz does not say anything about the predicted similarities or differences between these two groups for non-UG-governed language properties or how the latter may interact with aspects of language which are constrained by UG. By extending the cross-learner comparison to include L1 children in addition to L2 children and L2 adults, it will be shown in Chapter 3 and later in Chapter 6 that the interaction between the various UG- and non-UG-constrained components of language, that is between syntax, semantics and discourse/pragmatics, plays a crucial role in the acquisition of the property of language under investigation here, that is, direct object scrambling in Dutch.

It is important to note that translating the central hypothesis in Schwartz's approach into an experimental hypothesis is not entirely straightforward. It is generally the case that the hypothesis which states equality between two groups is the experimental null hypothesis, but here, this hypothesis, namely that the child L2 and adult L2 developmental sequences should be the same, is the hypothesis which is predicted by Schwartz's theory. If we find evidence to reject this hypothesis, that is, if there is evidence which strongly suggests that these two groups pass through *different* developmental sequences, then this hypothesis must be rejected. If the L2 children and L2 adults appear to pass through the *same* developmental sequences, then there will be no reason to reject this hypothesis. The problem here is that the failure to find enough evidence to reject an experimental null hypothesis could always be the result of two possibilities, either that this hypothesis is correct or that there *is* a difference between the two groups, but it was not found, that is, a Type II error may have been committed. The reason for pointing this out is not to show that this hypothesis cannot be tested, however, rather to spell out this difficulty and to highlight that the consequences which follow from this should be borne in mind in the interpretation of any results.

To summarise: This section has outlined some of the ways in which comparing L2 children and L2 adults can inform adult L2 acquisition research. Such a comparison is clearly crucial to the debate about maturational effects in L2 acquisition and it is this comparison which will be carried out here. More specifically, by investigating the extent to which these two groups are similar in their development in the target language, it has been argued that a child L2 ~ adult L2 comparison can arbitrate between UG-based and alternative, 'problem-solving' approaches to adult L2 acquisition.

In the following section, we consider how child L2 acquisition may be informative with respect to L1 acquisition. Similarly to the preceding discussion on adult L2 acquisition, this section focuses on the logic of the child L2 ~ child L1 comparison. Studies which make such comparisons will be reviewed in §1.5.

1.2.2 What child L2 can tell us about child L1 …

There are (at least) two ways in which the investigation of child L2 acquisition is illuminating with respect to child L1 acquisition: firstly, it can inform the debate about whether there is a critical period in L1 acquisition and secondly, it can serve as a means to evaluate theories of L1 acquisition.[11]

Testing for a critical period in L1 acquisition is restricted to the (fortunately) relatively few cases where exposure to primary language was unavailable during the early years of life (see e.g. Fromkin, Krashen, Curtiss, Rigley and Rigley (1974) on Genie, Lane (1976) on the 'wild boy of Aveyron' and Emmorey (2002) on deaf learners). As noted in the preceding section, child L2 data are often used as evidence for (and sometimes against) a critical period in L2 acquisition. If it transpires that there is no critical period for L2 acquisition, this will have no consequences for the question of whether there is a critical period in L1 acquisition. If, however, it turns out that there *is* a critical period in L2 acquisition, it then follows that there must also be a critical period in L1 acquisition. In this sense, then, child L2 data can inform child L1 acquisition research.

The second way in which child L2 acquisition can inform child L1 acquisition is by providing a useful testing ground for L1 acquisition theories. The child L2 ~ child L1 comparison has been particularly fruitful in assessing approaches to L1 acquisition which involve some element of maturation. Specifically, several researchers (e.g. Haznedar and Schwartz 1997; Henry and Tangney 1996; Ionin and Wexler 2002; Park 2000; Prévost 2003; Schwartz 2004) have argued that developmental patterns in L1 acquisition which are explained by making recourse to maturation should not be found in child L2 acquisition when the L2 children are beyond the relevant maturational point. The area which has generated the most research using this logic is the phenomenon of *Optional* or *Root Infinitives*. This will be discussed in §1.5.3.1.

To summarise: Child L2 data can be used to inform our understanding of L1 acquisition in two ways, as part of the critical period debate and as a testing

[11] Child L1 acquisition is used throughout this dissertation to refer to monolingual acquisition from birth.

ground for L1 acquisition theories. It is in the latter sense that child L2 data will be used in the present study.

1.3 Previous child L2 studies: An introduction

After having been the focus of an active research programme in the 1970s (e.g. Cancino, Rosansky and Schumann 1978; Ervin-Tripp 1978; Felix 1977; Hakuta 1976; Ravem 1978; Wode 1978, amongst many others), child L2 acquisition has somewhat taken the back seat in the by-now well-established field of generative L2 acquisition (see White 2003 for a recent overview). In the 1980s and early 1990s, the focus was mainly on adult L2ers. In recent years, however, child L2 acquisition has been the subject of renewed interest in the field. The remainder of this chapter presents an overview of recent research on this topic. An extensive table detailing the relevant literature, some of which is critically reviewed in §1.4 and §1.5, is provided in Appendix A. This table provides details of the subjects, including their L1, ages at first exposure and at time of testing, the method(s) used to collect the data, and the target language property under investigation. It includes a total of 49 different studies.

In earlier work, the driving force behind (child) L2 research was the question of whether L2 acquisition involves the same process as L1 acquisition. As the discussion in the preceding section indicated and as we shall see in the literature review which follows, the L2=L1 issue remains the focus of much of the research on child L2 acquisition today. Whereas earlier studies on child L2 acquisition were largely descriptive in nature and therefore lacking in explanatory adequacy (Lakshmanan 1995), recent work has sought to combine the insights in theoretical linguistics and language acquisition (L1 and L2) theory in order to deepen our understanding of both the L1 and L2 acquisition process.

The remainder of this chapter is divided into two parts. The first part (§1.4) reviews studies which compare child L2 acquisition with adult L2 acquisition and the second part (§1.5) reviews research on the child L2 ~ child L1 comparison. Given that most of the pre-1995 studies are included in Lakshmanan's (1995) review of generative child L2 syntax studies, the present discussion will predominantly – though not exclusively – focus on studies from the past ten years. The studies which are selected illustrate the state of current research on child L2 acquisition. Topics which have generated considerable interest include whether there is L1 transfer in child L2 acquisition and whether L2 children pass through a so-called Root Infinitive stage. In line with the approach taken in this thesis, particular attention is paid to those studies which use experimental rather than spontaneous production data and those which compare L2 children (when children) with L2 adults. Methodological issues pertinent to such a cross-learner comparison which these studies raise will be dealt with along the way.

1.4 Child L2 ~ Adult L2 : Differences and similarities

The child L2 ~ adult L2 comparison will be discussed in light of three issues: age, L1 transfer and developmental sequences.

1.4.1 Age (at time of testing)

The age of any L2er can be calculated either for the point of first exposure to the target language (TL) and/or for the time at which data collection takes place. With respect to age at first exposure, child L2ers differ from adult L2ers by definition. With respect to age at time of testing, child L2ers will differ in terms of age from adult L2ers if they are tested when they are still children, whereas they will potentially not differ from adult L2ers if they are tested when they are adults. The decision about whether to test child L2ers as children or adults largely depends on the purpose of the investigation: if this is to determine whether child L2ers become nativelike, then it would be better to test child L2ers when they are adults, whereas if the focus is on development, as is the case here, it is crucial to collect data from the child L2ers when they are still children, that is, while they are (still) developing.

1.4.1.1 Taking the L1 age of acquisition into account

Another important factor relating to age at time of testing, one which will have an effect on how the child L2 and adult L2 groups are determined, is the age at which the target language property in question (and its equivalent in the L1, if this exists) are acquired.[12] For most aspects of the target language, the situation is clear. Acquisition takes place before the age of four and the comparison between the L2 children and L2 adults will be as straightforward as the discussion in §1.2.1.2 implies. For properties which are acquired later than this age, however, the situation is more complicated. Imagine a target language property, P, which is not acquired by L1 children until around age ten. Imagine further that the L2ers' L1 either does not instantiate this property or does not have a directly transferable equivalent, or that if there is an equivalent, it, too, is not acquired until around age of ten. Now if the acquisition of P is maturationally determined, that is, if it is in some way crucially linked to age ten, it would be unreasonable to expect that L2ers who are younger than ten at time of testing should acquire P. This means that the child L2 group would have to be divided into sub-groups, those children older than age ten at time of testing and those younger than age ten at time of testing. Recall that on the definition employed here, L2ers who are first exposed to the target language at age eight or older are classed as adults. Consequently, they, too, would have to be divided into sub-groups, those older than and those younger than age ten at time of testing. In such a situation, therefore, a sub-group of the L2 adults

---

[12] Admittedly, establishing whether there is an L1 equivalent and what this should be is not always an easy task.

should actually be grouped with (a sub-group of) the L2 children and these will be expected to differ from the remaining L2 adults. This is the type of situation which will be encountered in Chapter 6.

1.4.1.2 Methodological implications

As noted above, most studies of the age issue in L2 acquisition are concerned with ultimate attainment and consequently they test child L2ers when they are adults. There is, however, an increasing number of studies which have collected data from child L2ers when they are children and which have compared these to adult L2 data. One of the earlier studies in this vein is Snow and Hoefnagel-Höhle (1982). This will briefly be discussed here in order to make a methodological point regarding the child L2 ~ adult L2 comparison.

Snow and Hoefnagel-Höhle's (1982) L1 English/L2 Dutch investigation employed a variety of production and comprehension tasks (see Appendix A for details) to test the L2 proficiency of a total of 81 children and adults whose age at time of testing ranged from three years to adults. The 51 beginners were tested on three occasions, at 4- to 5-month intervals. Briefly, it was found that on the first testing occasion, there were significant differences between the different age groups, with older subjects (12-15 and adults) producing more targetlike responses than the younger groups. These age-based differences remained on the second and third testing occasions, although they were considerably smaller. By this time, some of the younger children had surpassed the adults on several tests, but on the whole, the youngest child L2 group, the 3-5-year-olds, were consistently worse than the older groups. Thus, the authors conclude, older L2ers have the advantage over younger L2ers. The import of this much-cited conclusion, essentially based on rate data, regarding the age issue in L2 acquisition is debatable (see fn. 7). What is even more problematic, however, is the fact that the results might partly be a consequence of some of the testing methods employed.

In their battery of tests, Snow and Hoefnagel-Höhle included tasks which arguably favoured the older subjects. For example, in a sentence repetition task and a sentence translation task, subjects were asked to either imitate or translate sentences of increasing length (two to ten words in the repetition task) and grammatical complexity, which were read aloud to them. It is quite possible that the older subjects' superior performance on this task resulted from their relative cognitive maturity (e.g. greater working memory capacity) rather than their higher L2 proficiency. Furthermore, according to the information provided in the paper, subjects were tested on all (nine) tasks in one session, lasting 1.5 hours. It is inevitable that fatigue (and quite possibly boredom) influenced the results for (at least) the younger subjects.

The problems with Snow and Hoefnagel-Höhle's study highlight an important methodological point which should be taken into consideration for any study testing child and adult L2ers when the child L2ers are children at the time of testing: the chosen experimental methods should not be too cognitively challenging for the younger subjects, nor, on the other hand, should they be too easy for the older subjects. In addition to differences in cognitive capacities, such as working memory, the child and adult groups may also differ in the amount of metalinguistic knowledge they have. Older subjects, if they have had language instruction and perhaps even if they have not, will have metalinguistic knowledge which they could potentially employ in certain data collection procedures.[13] Striking a balance between L2 children and adults in these terms, that is, trying to ensure that one group is not favoured over the other, is a challenge. There are, however, two possible ways in which this challenge could be countered. The first option would be to tailor a particular task to the different (cognitive) abilities of each age group, i.e. use different versions of the task with the different age groups. Although this would address the problem, using different tasks with the different age groups (at least without careful piloting and validation) could make results difficult to interpret as it introduces another variable to the comparison. Disparate outcomes for the different age groups could potentially be due to the differences in the task as well as any differences in the subjects' interlanguage grammars. Another option would be to ensure that the target language property which is being tested is sufficiently sophisticated to ensure that metalinguistic knowledge cannot be applied. If this is the case, using an 'easy task' with L2 adults will be less problematic. This will be the strategy pursued in the present study.

1.4.1.3 Summary

This section has outlined two ways in which age at time of testing is pertinent to the child L2 ~ adult L2 comparison. Firstly, depending on the aspect of linguistic development under examination, L2 children and adults may or may not belong to the same 'age group'. Secondly, the different cognitive capacities inherent in different age groups have consequences for the type of method which can be used to compare these two groups.

1.4.2 L1 transfer and the initial state

The second point of consideration regarding the child L2 ~ adult L2 comparison is L1 transfer and the related topic of the initial state in L2 acquisition. Possibly the most obvious difference between L2 children and L2 adults, on the one hand, and L1 children, on the other, is the fact that the L2 children and adults already have knowledge of another language. Consequently, their interlanguage grammar may make use of this knowledge; in other words, there is the potential for L1 transfer. Although this might, to a certain extent, seem like stating the obvious, it is only relatively recently that the issue of L1

---

[13] Differences between older and younger subjects in terms of metalinguistic knowledge/instruction are a problem for Rounds and Kanagy's (1998) study, for example.

transfer in child L2 acquisition has been systematically investigated. This section details the outcomes of some of these investigations.

1.4.2.1 Haznedar (1997)

Perhaps some of the most convincing evidence for L1 transfer in child L2 acquisition comes from Haznedar's (1997) study on the L2 acquisition of English by a Turkish-speaking child, Erdem. For the first two months of his stay, Erdem, was mainly at home with his Turkish-speaking parents. In the third month in the UK, aged 4;3, he started attending an English-speaking nursery school. It was after just one month of exposure in the nursery that data collection began. Erdem's data thus represent the very early stages of the L2 acquisition process. Haznedar examines the acquisition of the VP. Erdem's L1, Turkish, is a head-final language with an SOV word order in both main and embedded clauses (cf. the TL, English, which is SVO). Utterances containing a verb were classified according to whether they exhibited a VX (head-initial) or XV (head-final) order, where X stands for any other VP-material, such as an adverbial or direct object. A clear pattern is observed: during the first nine samples, spanning approximately the first two-and-a-half months of data collection, Erdem's verbal utterances were predominantly XV, as in (5), whereas after this point, the VX order, illustrated in (6), prevailed. Specifically, until sample 9, 91.3% (21/23) of Erdem's utterances had an XV order, whereas from sample 9 onwards, just 1.7% (14/806) of his utterances had this order.

(5) a. I something **eating** (Erdem, sample 8)
b. Something **playing** # dinosaur **playing**
c. This cartoon television **looking**
(Haznedar 1997:248, ex.249)

(6) a. **Going** this way (Erdem, sample 10)
b. This teddy bear **is looking** that
c. I'**m going** to playground (Erdem, sample 11)
(Haznedar 1997:248, ex.210)

Erdem's data are convincing evidence that there is L1 transfer in child L2 acquisition. (They also demonstrate that data collection has to begin very early if such initial stages in the acquisition process are to be captured.) Haznedar's detailed case study provides very valuable information about the initial stages and subsequent development in child L2 acquisition. In order to investigate the role of L1 transfer in more detail, however, it is necessary to compare L2 children with different L1s in their acquisition of the same L2. The following two sections review studies which do exactly that.

1.4.2.2 Haberzettl (1999)

Haberzettl (1999) examines child L2 German. Given that several of the studies which are reviewed in the remainder of this chapter deal with the acquisition of

German (or Dutch, which behaves similarly in all crucial respects), the basic word order properties of this language will now be outlined briefly.

German is a V2 language with SOV in embedded clauses. OV is traditionally assumed to be the underlying word order, that is, the VP is assumed to be head-final. Thus, when a finite and non-finite verb appear together in a main clause, the non-finite verb appears clause-finally, to the right of any VP-internal material, as in (7)-a (cf. (7)-b). When a non-subject constituent is in initial position, as in (8)-a, subject-verb inversion ensues, such that the finite verb appears in second constituent position (cf. (8)-b, where the finite verb is in third constituent position).

(7) a. Rick **hat** gestern den Garten **umgegraben**
Rick has yesterday the garden up-dug
'Rick dug up the garden yesterday.'

b. *Rick **hat** **umgegraben** den Garten gestern
Rick has up-dug the garden yesterday

(8) a. Gestern **hat** Rick den Garten umgegraben
yesterday has Rick the garden up-dug
'Yesterday, Rick dug up the garden.'

b. *Gestern Rick **hat** den Garten umgegraben
yesterday Rick has the garden up-dug

Haberzettl reports on two L1 Turkish children acquiring German, aged six years, and compares them with an L1 Russian child acquiring German, aged eight years. At the start of data collection, the Turkish children had had about six months' exposure and the Russian child about three months.[14] As mentioned above, Turkish is an SOV language with this order occurring in both main and embedded clauses. Russian, on the other hand, is SVO. Although the lack of raw figures means that it is difficult to interpret the robustness of the patterns in these data, the general tendency is clear: for the least advanced Turkish-speaking L2er, early verbal utterances have the order XV, as in (9), whereas those of the Russian-speaking child are VX.[15]

(9) a. Das [= das Kind] Hotel **mach** [= geht zum Hotel] (Ne, 9 months' exposure)
that [= the child] hotel make [= goes to the hotel]

b. Das Kind […] Eis **kaufen**
the child ice-cream buy$_{\text{INF}}$

(Haberzettl 1999:163)

[14] Depending on her exact age at arrival (which is not given), the Russian-speaking child might be classed as an L2 child or an L2 adult on the definitions used here.

[15] The article from which these examples are taken is in German; without any contextual information, it is difficult to provide an accurate translation.

For the Russian-speaking L2er in her third month of exposure, 80% of utterances containing a finite and non-finite verb, a subject and some other material (such that the position of the two verbs relative to each other could be established) contained the finite and non-finite verb directly adjacent to each other, as in (10). This proportion is considerably reduced in the eighth month of exposure (58%) and by the tenth month, it is virtually non-existent (3%).

(10)	Der Räuber	__	will	**schießen**	den __	(An, 3 months' exposure)
	the robber		wants	shoot	the	
						(Haberzettl 1999:160)

Although the utterances produced by the Turkish-speaking child in (9) are in fact indistinguishable from a typical utterance produced by an L1 German child in the RI stage and could therefore indicate that L2 children pass through this stage in much the same way as L1 children do, a comparison of these utterances with the data from the Russian-speaking child suggests that L1 transfer is a more likely explanation.

1.4.2.3 Whong-Barr and Schwartz (2002)

Whong-Barr and Schwartz (2002) also examine the issue of transfer by comparing child L2ers with different L1s. They investigate the acquisition of the dative alternation in English by Korean-speaking (n=5) and Japanese-speaking children (n=5). The Japanese-speaking children were first exposed to English between the ages of 4;2 and 7;8 and they were between 7;3 and 8;11 at time of testing. The Korean-speaking children were between 6;11 and 10;2 at time of testing and between 3;7 and 8;1 at first exposure. In English, certain ditransitive verbs, such as *give*, allow their arguments to be realised in two different syntactic environments: as prepositional datives, as in (11)-a through (14)-a, and as double object datives, as in (11)-b through (14)-b. As the pairs in (12) and (14) illustrate, however, not all verbs allow this alternation.

(11) Alternating *to*-dative
a. The sheep threw the spoon to the giraffe (Prepositional *to*-dative)
b. The sheep threw the giraffe the spoon (Double-object *to*-dative)

(12) Non-alternating *to*-dative
a. The pig said the answer to the giraffe (Prepositional *to*-dative)
b. *The pig said the giraffe the answer (Double-object *to*-dative)

(13) Alternating *for*-dative
a. The giraffe made a cup of tea for the pig (Prepositional *for*-dative)
b. The giraffe made the pig a cup of tea (Double-object *for*-dative)

(14) Non-alternating *for*-dative
a. The tiger held the money for the sheep (Prepositional *for*-dative)
b. *The tiger held the sheep the money (Double-object *for*-dative)

The acquisition of the dative alternation presents a poverty-of-the-stimulus problem: the child must come to know which verbs allow both forms and more importantly, which do not, on the basis of the limited, positive evidence alone. The L2 children's L1s, Japanese and Korean, differ with respect to whether they allow double object constructions. In Japanese, double object constructions are ungrammatical and hence there is nothing which is analogous to the dative alternation in English. Korean, on the other hand, has two forms of the double-object construction, both of which can be considered equivalent to the benefactive-*for* forms given in (13) and (14). These are used with a wider range of verbs than the English dative alternation, however. The first alternating form in Korean involves the benefactive verbal morpheme *cwu-*, which, as exemplified in (15), is obligatory (cf. (15)-c). The second, which is similar to the prepositional dative form in English ((13)-a and (14)-a), involves the adverbaliser *-ulwihay*, which marks the object as the goal/recipient. This is exemplified in (16).

(15)	a.	Mia-ka	Yong-eykey	kulim-ul	kuly-e	**cwu**-ess-ta
		Mia-NOM	Yong-for/DAT	picture-ACC	draw-L	BEN-PAST-DECL
		'Mia drew a picture for Yong.'				
	b.	Mia-ka	Yong-ul	kulim-ul	kuly-e	**cwu**-ess-ta
		Mia-NOM	Yong-ACC	picture-ACC	draw-L	BEN-PAST-DECL
		'Mia drew Yong a picture.'				
	c.	*Mia-ka	Yong-eykey/ul	kulim-ul	kuly-ess-ta	
		Mia-NOM	Yong-for/DAT/-ACC	picture-ACC	draw-PAST-DECL	

(16)	Mia-ka	Yong-**ulwihay**	kulim-ul	kuly-ess-ta
	Mia-NOM	Yong-for	picture-ACC	draw-PAST-DECL
	'Mia drew a picture for Yong.'			

An oral grammaticality judgment task was used to elicit judgements on sentences containing the four classes of verbs illustrated in (11) through (14). Sentences were presented as minimal (grammatical/ungrammatical) pairs using a context set up by means of props.

The results were as follows. All groups (Korean, Japanese, native controls) were targetlike on the grammatical prepositional dative sentences ((11)-a through (14)-a). On the double-object *to*-datives, the English native-speaker control children showed evidence of overgeneralisation, accepting the illicit form given in (12)-b in 37.9% (11/29) of items. The L2 children also accepted these illicit forms at an average rate of 60% (12/20) for the Japanese-speaking children and 40.9% (9/22) for the Korean-speaking children. It is generally the more proficient subjects who make this overgeneralisation.[16] The situation is rather different for the double-object *for*-datives, however. Here, the native-speaker children always correctly accept sentences such as (13)-b, but they virtually never incorrectly accept sentences such as (14)-b. The Japanese-speaking children accept both licit and illicit forms at comparable rates, whereas the Korean-speaking children make the targetlike distinction, generally accepting the licit sentences and accepting the illicit ones much less frequently.[17] In other words, the Japanese – but not the Koreans – overgeneralise the double-object *for*-datives in the same way as they overgeneralise the double-object *to*-datives. The difference between the two child L2 groups is statistically significant (ANOVA (one-way): $F = 8.225$, $p = .021$). As the authors note, it is unlikely that this difference is due to distinct proficiency levels for the two groups because the ranges of proficiency scores in each group are comparable. Rather, Whong-Barr and Schwartz (2002) claim, this difference is a consequence of L1 transfer: the only sentence-type which elicits different responses for the two groups, namely benefactive *for*-datives, corresponds to where the L1s differ. Recall that Korean has a construction which is analogous to the dative alternation, using the benefactive verbal morpheme, *cwu-*, whereas Japanese does not have any equivalent. This contrast is reflected in these L2 children's productions in English.

1.4.2.4 Other studies and summary

Another example of a study dealing with transfer in child L2 acquisition is Jin (2003). Jin also finds evidence for L1 transfer in terms of the extent to which certain types of errors are made. In his investigation of the acquisition of determiners in child L2 English, his Korean-speaking subjects (aged five to eight years) could more accurately imitate types of determiners which are available in the L1 (namely, possessives and demonstratives) than those which are not (such as articles).

All of the studies reviewed here examine L2 children (and sometimes L1 children) only. None provide comparative data from L2 adults. In fact, there are very few studies which directly compare L2 children with L2 adults (just 12 in total in the summary table in Appendix A) and of these, most do not systematically examine (or sometimes even control for) L1 transfer. Those studies which do compare the two groups of L2ers in these terms have found that they are similar to each other, that is, for L2 children and adults with the same L1, similar transfer effects have been observed and, consequently, these

---

[16] Proficiency was measured by means of an independent picture description task. For details, see Chapter 4. Whong-Barr and Schwartz's (2002) proficiency measure forms the starting point for the measure which will be used in the present study.

[17] The relevant figures are as follows: the Japanese-speaking children accept illicit and illicit forms at a rate of 61.1% (11/18) and 70% (14/20), respectively, and the Korean-speaking children accept licit sentences in 86.4% (19/22) of cases and illicit ones in 14.3% (3/21) of cases.

two groups differ from L1 children. For example, Van de Craats (2001) observes that in their acquisition of the Dutch possessive pronoun system, Turkish-speaking children and adults both make use of the same non-targetlike forms. Also, Gilkerson (2004), in her study of the acquisition of particle verbs in the L2 English of Spanish-speaking children and adults, finds that L2 groups make similar kinds of errors and these errors are not attested in L1 children. Her study will be discussed in more detail in the following section.

In summary, the studies reviewed in this section indicate that there is accumulating evidence that as in adult L2 acquisition, the start of the child L2 acquisition is characterised by L1 transfer.

1.4.3 Developmental sequences

In the third and final section on child L2 ~ adult L2 comparisons, our attention turns to developmental sequences. Schwartz (1992) proposes that in order to address the question of whether (adult) L2 acquisition involves the same processes as L1 acquisition, we need to compare the developmental sequences of adult L2ers with those of child L2ers (given the logic laid out in §1.2.1.2). First, I briefly review some of the older, morpheme order studies, before moving to the developmental sequence data discussed in Schwartz's original paper. The bulk of this section consists of a detailed review of some of the more recent child L2 ~ adult L2 studies which have examined developmental sequences.

1.4.3.1 Earlier studies

Probably the most well-known developmental sequence data come from the grammatical morpheme order studies carried out in the 1970s. For example, following work in L1 acquisition by Brown (1973) and De Villiers and De Villiers (1973), Dulay and Burt (1973) calculated a 'difficulty order' for eight morphemes in child L2 English. Using elicited speech from 151 Spanish-speaking children acquiring English, they calculated a morpheme (functor) ratio based on the average number of times a given morpheme was supplied in obligatory contexts. The following morphemes were examined: present progressive *–ing*, plural *–s*, the irregular past, possessive, articles, third person singular *–s*, the contractible copula and the contractible auxiliary. The L2 children, who were taken from three different sociolinguistic contexts where different amounts of exposure to English were available, were found to perform similarly on all eight morphemes. The relative accuracy on each morpheme was interpreted as the order in which these morphemes were acquired (most accurate < least accurate). In a follow-up study, Dulay and Burt (1974) found that the order of acquisition of these morphemes for Chinese-speaking child L2ers was comparable to that of the Spanish-speaking child L2ers. Furthermore, in a similar study of L2 adults, Bailey, Madden and Krashen (1974) found that the relative accuracies for these morphemes for

L2ers with a variety of L1s (Spanish, Greek, Persian, Italian, Turkish, Japanese, Chinese, Thai, Afghan, Hebrew, Arabic and Vietnamese) were comparable to those in Dulay and Burt's (1973) study. These authors suggest that the results indicate that there might be a 'natural sequence' to second language acquisition.

Both the results and the claims based on the results of the morpheme order studies have been challenged, however. Larsen-Freeman and Long (1991) summarise the objections which have been expressed. First, despite the observed cross-group similarities, significant differences also existed. Second, different subsets of grammatical morphemes were often tested with the result that the overlap between cross-learner comparisons is minimal. Finally, the group of (free/bound, NP/VP) morphemes is rather heterogeneous and hence it is not clear in what respect an acquisition order (common or not) is informative with respect to the L2 acquisition process. It is for this latter reason, namely that such 'difficulty order' studies do not show the development of a particular phenomenon, that Schwartz (1992) excludes them from her review of the then available literature, to which we now turn.

Cancino, Rosansky and Schumann (1978) examine the acquisition of negation and interrogatives in the longitudinal spontaneous data from six Spanish-speaking L2ers of English: two children (aged four and five at time of testing), two adolescents (aged ten and 12 at time of testing) and two adults (aged 25 and 30 at time of testing). They find that members of all three groups, who had been exposed to English for between one and four months at the start of data collection, (more or less) pass through the same developmental stages. The stages for negation are given below:

(17)	Stage I:	neg + verb
	Stage II:	(unanalysed) *don't* + verb/aux
	Stage III:	aux + neg
	Stage IV :	(analysed) *don't*

Schwartz argues that the similarity between these groups suggests that child L2 acquisition and adult L2 acquisition make use of the same language acquisition device and, given the logic laid out above (cf. §1.2.1.2), this must be UG.

Similarities in L2 child and L2 adult developmental sequences are also observed in the L2 acquisition of German word order by speakers of Romance (Spanish/Italian/Portuguese) languages. Comparing longitudinal data taken from three L1 Italian children (age eight years at time of testing - Pienemann 1980; 1981), longitudinal data from one L1 Portuguese and two L1 Spanish adults (Clahsen 1984) and cross-sectional data from 45 L1 Romance adults (Meisel, Clahsen and Pienemann 1981), Schwartz (1992) notes that both L2 children and adults pass through the same developmental stages:

(18)	Stage I:	S $V_{[+finite]}$ ($V_{[-finite]}$) O	
	Stage II:	(Adv/PP) S $V_{[+finite]}$ O	
	Stage III:	S $V_{[+finite]}$ O $V_{[-finite]}$	Particles, participles, infinitives in clause-final position
	Stage IV :	XP $V_{[+finite]}$ S …	Verb-second (V2/'subject inversion')
	Stage V:	S $V_{[+finite]}$ (Adv) O	
	Stage VI:	…, *daß* 'that' S O $V_{[+finite]}$	Distinction made between root and embedded clauses

(Schwartz 1992:11, ex. 12)

Again, given the assumptions which she makes, these similarities constitute evidence for the claim that both adult and child L2ers make use of UG.

Given that there were very few data available at the time, the data which Schwartz (1992) reviews are rather limited. The number of subjects is rather low at just two per group in the Cancino *et al.* paper and only three child L2ers in the German word order data. Furthermore, although, as Schwartz points out, Italian, Portuguese and Spanish are similar with respect to the relevant word order properties, a cleaner comparison would hold the L1 constant. Finally, all the data are spontaneous production data. Although these data also have the advantage of being longitudinal, which means that developmental sequences can be detected more easily and more importantly, they can be observed in one and the same learner, spontaneous production data are less than ideal because they may seriously underestimate an L2er's linguistic competence (because of avoidance strategies, for example).

A number of recent studies comparing child L2 acquisition with adult L2 acquisition go some way toward redressing this gap, using elicited production tasks to target the L2 property of interest. Each of these studies is reviewed in turn. Subsequently, another recent paper, Dimroth (2005), which uses longitudinal spontaneous production data to investigate the acquisition of finiteness and negation in German, rounds up this section.

1.4.3.2 Gilkerson (2004)

Gilkerson's (2004) study is based directly on Schwartz's (1992) proposal. She compares Spanish-speaking children (n=32) with Spanish-speaking adults (n=33) in their acquisition of particle verbs in English. The age of first exposure was around five years for the children and 18 years or older for the adults. At time of testing, the adults were between 21 and 45 years and the children between 4;4 and 8;7.

The particle in particle verbs can appear either directly adjacent to the verb, as in (19)-a, or separated from the verb by the direct object, as in (19)-b. Following Wurmbrand (2000), Gilkerson analyses particle verbs such as those in (19) – so-called transparent particle verbs – as part of a small clause (= AgrP) structure. As the structures in (19)-a' and (19)-b' indicate, the split order is the base structure and the adjacent order is derived from this.

(19) a. She **threw out** the garbage (adjacent particle verb)

a'. …$[_{vP}$ [**throw out**$_i]_j$ $[_{VP}$ t_j $[_{AgrP}$ $[_{DP}$ the garbage $[_{Agr}$ t_i $[_{PrtP}$ $t_i]]]]]]$

b. She **threw** the garbage **out** (split particle verb)

b'. …$[_{vP}$ **throw**$_i$ $[_{VP}$ t_i $[_{AgrP}$ $[_{DP}$ the garbage $[_{PrtP}$ **out**$]]]]]$

The meaning denoted by particle verbs in English is expressed in Spanish, the L1, using simple verbs. An example is provided in (20).

(20) Sacó la basura
threw out the garbage
'S/he threw out the garbage / S/he threw the garbage out.'

In order to elicit particle verbs, an experimenter acted out a number of actions for the subject who in turn had to describe them to another experimenter who was blindfolded. Fourteen particle verbs were elicited, twice each. In addition to the elicitation experiment, a second task, a modified version of the Curtiss and Yamada *Comprehensive Language Evaluation – Elicited Production* (1985), was carried out in order to assess the subjects' L2 proficiency. On the basis of the results of this normalised English proficiency test, subjects were divided into three levels. These levels were used to infer development from the cross-sectional data. There were 11 children and 11 adults in each level (except Level 2, where there were ten children).

The results are presented in Table 1. For each group (child/adult) and level (1/2/3), the table gives the proportion of responses where the particle was either in split or adjacent position or omitted altogether, out of the total number of utterances for these three categories.[18]

[18] Subjects also produced responses with a PP instead of the particle or a simple verb, both of which do not provide relevant information with respect to particle verbs, or they produced something else altogether (e.g. they omitted the verb or the object). The percentages in Table 3 are based on Figures 1 and 2 in Gilkerson's paper.

Table 1. Gilkerson (2004): Production of particle verbs in L2 children and L2 adults

Response Type	L2ers						Native adults	
	Level 1		Level 2		Level 3			
	N	%	N	%	N	%	N	%
Children								
Particle omission	18	31.05%	9	11.4%	4	5.0%	1	1.4%
Split	22	37.9%	53	67.1%	35	44.3%	27	39.7%
Adjacent	18	31.05%	17	21.5%	40	50.6%	40	58.8%
Total	58		79		79		68	
Adults								
Particle omission	11	22.4%	12	20.0%	7	12.1%	0	0%
Split	1	2.0%	0	0%	8	13.8%	24	28.9%
Adjacent	37	75.5%	48	80.0%	43	74.1%	59	71.1%
Total	49		60		58		83	

There are both similarities and differences between the L2 children and adults. The two groups are similar, Gilkerson claims, in that a decrease in particle omission is coupled with an increase in split particle verbs. This occurs at the transition from Level 1 to Level 2 for the children and from Level 2 to Level 3 for the adults. Gilkerson takes the increase in the split particle verbs coupled with a decrease in particle omission to indicate the acquisition of the underlying structure of these verbs (cf. (19)-b'). The two groups differ in that the children are targetlike in particle omission by Level 3, whereas none of the adult groups are (completely). Furthermore, when the L2 adults produce particles, they appear most frequently in adjacent orders, for all three levels. This contrasts with the L2 children, who produce the split orders at Level 1 already. Gilkerson claims that this is because the children are misanalysing the particle as an adverb which projects as an AdvP, as in the L1 Spanish, resulting in the split order with an underlying structure as follows:

(21) …[$_{\text{VP}}$ [$_{\text{V}}$ pulled [$_{\text{AdvP}}$ [$_{\text{DP}}$ her pants [$_{\text{Adv}}$ up]]]]] (adverb analysis)

Gilkerson notes that L1 transfer could also result in the adjacent order and/or particle omission. An adjacent order would result if the particle verb were analysed as a single complex verb, as in the Spanish simplex equivalents, and particle omission would for example follow from the transferred interlanguage grammar not being able to accommodate it. The structures for both of these options are given in (22)-a and (22)-b, respectively.

(22) a. …[$_{\text{VP}}$ [$_{\text{V'}}$ [$_{\text{V}}$ [pulled up][$_{\text{DP}}$ her pants]]]] (complex verb analysis)
b. …[$_{\text{VP}}$ [$_{\text{V'}}$ [$_{\text{V}}$ [pulled][$_{\text{DP}}$ her pants]]]] (particle omission)

The fact that both split and adjacent orders are predicted by transfer as well as being possible orders in the TL means that it is very difficult to tease apart productions which result from successful acquisition from those which result from transfer. The concomitant decrease in particle omission and rise in split orders which Gilkerson notes is indeed suggestive of some change in the interlanguage grammar in the direction of the targetlike grammar, but the lack of significant differences between Levels 2 and 3 for the adults on either of these variables weakens the claim for similarity between the child L2 and adult L2 groups in these terms.

To summarise: Gilkerson finds that although there are some differences between the two groups in the acquisition of particle verbs in English, Spanish-speaking children and Spanish-speaking adults pass through similar developmental stages. Teasing apart targetlike behaviour and L1 transfer for these learners was found to be problematic, however. Despite these problems, however, Gilkerson's study is exemplary in terms of eliciting data using the same task to compare L2 children with L2 adults. Furthermore, the use of an independent means of assigning subjects to different proficiency levels serves to validate the cross-group comparisons which she makes.

1.4.3.3 Weerman, Bisschop and Punt (2003)

The second child L2 ~ adult L2 comparison study reviewed in this section, that of Weerman, Bisschop and Punt (2003) (see also Weerman 2002), concerns the acquisition of adjectival inflection in Dutch.[19] The L2 adult subjects (n=14) were first exposed to the L2 between the ages of 14 and 59 and they were between 16 and 61 at time of testing. Unlike in Gilkerson's study, the L2 children here are no longer children at the time of testing, that is, after being exposed to Dutch at age four years, these subjects were aged between 16 and 18 at time of testing. It will be shown that this contributes to one of several methodological problems which undermine the results of this study.

Whether an adjective is inflected or not depends on its status as either attributive or predicative, and if attributive, on the type of noun it modifies. Attributive adjectives are inflected with *–e,* pronounced as schwa, as in (23)-a to (23)-d, except when modifying a singular, indefinite, neuter noun, as in (23)-e. Predicative adjectives, as in (23)-f, are never inflected.

(23) a. Dit is een **lekkere** peer (singular, indefinite, non-neuter)
this is a delicious pear
'This is a delicious pear.'

b. Dit zijn **lekkere** peren (plural, indefinite, non-neuter)
these are delicious pears
'These are delicious pears.'

[19] The L2 data presented here are taken from Punt (1998).

c. Dit is het **lekkere** appeltje (singular, definite, neuter)
this is the delicious apple
'This is the delicious apple.'

d. Dit zijn **lekkere** appeltjes (plural, indefinite, neuter)
these are delicious apples
'These are delicious apples.'

e. Dit is een **lekker** appeltje (singular, indefinite, neuter)
this is a delicious apple
'This is a delicious apple.'

f. De peer / het appeltje is **lekker** (predicative)
the pear the apple is delicious
'The pear/the apple is delicious.'

Weerman *et al.* (2003) used a series of elicited production tasks to ascertain whether L1 children, L2 children and L2 adults were able to produce the inflected form in the correct contexts only. The first task determined whether the subjects had targetlike knowledge of the grammatical gender of certain Dutch nouns and of the indefinite determiner by using pictures and puppets to elicit determiner+noun combinations for definite and indefinite NPs. The second task tested whether subjects could combine the same nouns with targetlike attributive adjectives. In the definite condition, the subject was told a short story in the presence of a blindfolded puppet, who then asked a question about the story and in responding the subject was expected to produce a noun modified by a preceding adjective. In the indefinite condition, subjects were presented with two pictures and by means of a prompt, they were asked to spot the difference between two. Again, this elicited prenominal adjectival modification. Targetlike responses include adjectival inflection for all noun types except singular, indefinite, neuter nouns (such as (23)-e above). An example scenario, taken from Punt (1998), for the singular, indefinite, non-neuter condition is given in (24).

(24) Weerman *et al* (2003):
Example scenario for singular, indefinite, non-neuter condition
Experimenter: Ernie staat bij de garage
Ernie stands by the garage
'Ernie is stood by the garage.'

Hij ziet twee auto's
he sees two cars
'He sees two cars.'

Wat is het verschil?
what is the difference
'What is the difference?'

	Dit	is	een	…
	this	is	a	
	'This is a …'			
Subject:	…	kleine	auto	
		small	car	
	'…small car.'			

The third and final task was designed to elicit a predicative adjective as part of a completion of the experimenter's utterance about a picture. Again, the nouns used in this task were also tested in the first task.

The number of targetlike responses on the second task was calculated by taking responses on the first task into account, that is, the targetlikeness of adjectival inflection in combination with a particular noun in the second task was judged on the basis of the gender which subjects assigned to the same noun on the first task. For example, if a subject incorrectly combined the neuter noun *huis* 'house' with the non-neuter definite determiner *de* 'the' in the first task, it was assumed that this noun had non-neuter gender in the subject's interlanguage grammar and responses on the second task were assessed accordingly. Although this adjustment takes into account the possibility that the gender of a given noun in a subject's interlanguage grammar might not be the same as in the target language, because each noun was only tested once in the first task, it is unclear how reliable such an adjustment actually is. This problem is further confounded by the observation that interlanguage is prone to variability.[20]

Bearing this problem in mind, we turn to the results. When they produced predicative adjectives, all three groups always produce targetlike, uninflected forms. They vary, however, in their responses on attributive adjectival inflection. The main observation which Weerman *et al.* make concerns the singular, indefinite, neuter condition (cf. (23)-e), which, recall, is where the adjectival inflection, *-e,* is *not* required. The L2 children are considerably less targetlike on this condition than on the other conditions. Only two of the six are completely targetlike; as a group, they score 76.7% (105/137). The 32 non-targetlike responses in this context result from overgeneralisation of the *–e* ending. This pattern is also observed in the L1 children: the younger children (3-, 4- and 5-year-olds), who were tested using exactly the same task, also overgeneralise the *–e* ending in the indefinite, neuter condition, (incorrectly) producing it in 27.5% (81/295) of items. This contrasts with the older L1 children (6- and 7-year-olds), who are targetlike. The L2 children thus pattern like the L1 children in terms of error type, but not in terms of ultimate attainment.

The L2 adult data pattern differently from the L1 and L2 child data. There is much more variability in all conditions and there are both subjects

[20] See Romaine (2003) for an overview of variability in L2 acquisition.

who overgeneralise the *–e* form (3/14) in the singular, indefinite, neuter condition and subjects who overgeneralise the *–ø* form (7/14) in the other conditions (all requiring *–e*). The remaining subjects produce *–e* and *–ø* across conditions. Thus, Weerman *et al.* conclude that, in terms of types of errors, adult L2 acquisition is different from both child L1 acquisition and child L2 acquisition. Note that L1 transfer is assumed not to account for the variability found within the adult L2 groups because those L2ers whose L1s pattern similarly with respect to the relevant properties (e.g. presence/absence of definite/indefinite articles, adjectival inflection) do not produce similar responses on the experimental tasks (Punt 1998). For example, those (adult) L2ers whose L1 has an indefinite article (Berber, Danish, French, Kurdish, Moroccan Arabic, Persian and Turkish) do not score significantly better on the first test, where they are asked to produce an indefinite article, than those L2ers whose L1 ('Chinese', Hindi, Thai) does not have an indefinite article. The scores in the former group range from 0% to 87% and in the latter from 46% to 92%.

Despite the fact that this study should be applauded for using the same task to collect data from the three different groups of learners, it suffers from several methodological problems. First of all, the overlap in L1s across the child L2 and adult L2 groups is only partial, with just three languages in common to both (Berber, 'Chinese' and Turkish). Given that the other eight languages are not constant across children and adults, this constitutes another source of variability in the data. In order to systematically test for L1 transfer, more subjects with the same L1 are needed. Furthermore, even if, by holding the L1 constant across both groups, the role of L1 transfer was controlled for, there is another variable on which these two groups differ, namely length of exposure. In other words, there is a confound between length of exposure and a subject's status as L2 child or L2 adult. The L2 adults have had considerably less exposure to Dutch than the children. Irrespective of this problem, without any independent measure of proficiency, it is impossible to know whether the various response patterns which are observed in the experimental data are, for example, characteristic of a particular proficiency level. In other words, nothing at all can be said about development. More seriously, without some indication of proficiency level, it is questionable whether the L2 adults should be considered as one group at all. It is quite possible that the heterogeneity in their response patterns results from the heterogeneity in their proficiency levels, for example. Thus, any differences observed between the L2 children and the L2 adults could therefore be due to either one of these variables. For example, the observation that the responses produced by the L2 adults are more variable than those given by the L2 children could be due to (i) the different L1s found in the two groups, (ii) the fact that (at least some of) the L2 adults have not reached their end-state, and/or (iii) the two groups containing learners at different levels of proficiency. Some of these methodological problems are being addressed, however, in ongoing work by Weerman and colleagues in the 'Variation in Inflection' project at the University of Amsterdam. It is to one part of this project that we now turn.

1.4.3.4 Blom and Polišenská (to appear)

In a recent paper, Blom and Polišenská (see also Blom 2005) investigate the acquisition of the morphosyntactic properties of finiteness in L2 children and adults. More specifically, they consider whether these two groups are similar in their acquisition of verb form and verb placement and compare this with child L1 data taken from the literature (de Haan 1996; Zuckerman 2001).[21] They observe that in this domain, L2 children generally exhibit more similarities with L1 children than with L2 adults.

Blom and Polišenská's subjects are Turkish-speaking children (n=10) and adults (n=5) and Moroccan-Arabic – sometimes plus Tarafit – speaking children (n=17) and adults (n=6). The adults, aged between 22 and 44 years when tested, were first exposed to Dutch after the age of 15. The age of exposure for the children was around four, the age at which they started school, and they were between five and eight years old at the time of testing.[22] The relevant properties of Dutch regarding verb placement are identical to those summarised in §1.4.2.2 for German: Dutch is underlyingly OV and it is a V2 language. Consequently, in main clauses all finite verbs occur in second constituent position and non-finite verbs are realised in sentence-final position. In embedded clauses, finite verbs also occur in final position. The verbal inflectional paradigm in Dutch is relatively poor. There are just three different endings: *–ø* for 1SG, *–t* for 2SG and 3SG, and *–en* for all plural persons. The canonical word order in Turkish is SOV and in Moroccan Arabic and Tarafit, it is SVO. None of these languages has V2.

Blom and Polišenská used a number of cleverly designed production tasks to elicit different word orders (main clause vs. embedded clause) and different verbal agreement inflections. Subjects were also tested for proficiency using a sentence-repetition task taken from the standardised Dutch test for Turkish-, Moroccan-Arabic- and Surinamese-speaking children. Data from the various experimental tasks were collapsed and the following calculations were made: (i) percentage of correct verb placement, (ii) percentage of dummy auxiliaries, and (iii) substitution errors. L2 children and L2 adults were found to be similar with respect to some substitution errors, but they differ in terms of verb placement, use of dummy auxiliaries and other substitution errors. Each of these results is dealt with in turn.

For verb placement, the following calculations were made: percentage of utterances with the target VO word order in main clause contexts,

---

[21] See Polišenskà (2005) for related experimental work on L1 children.

[22] These children were born in The Netherlands and will therefore probably have had some exposure to Dutch before the age of four (E. Blom, personal communication, 23 June, 2005).

percentage of utterances with target OV word order in embedded clause contexts and percentage of subject-verb inversion in obligatory contexts (e.g. where an adverb appears in first constituent position). The results for both groups, plus comparative data from L1 children (n=10, 3;0-3;11) taken from Zuckerman (2001), are presented in Table 2.

Table 2. Blom (2005, Table 2): Percentage of correct verb placement for L1 children, L2 children and L2 adults (n gives number of tokens)

Group	**VO % correct**	**OV % correct**	**Inversion % correct**
Child L1	94% (n=145)	98% (n=141)	n/a
Child L2 (Turkish)	86% (n=148)	89% (n-152)	85% (n=148)
Child L2 (M. Arabic/Tarafit)	91% (n=301)	81% (n=311)	73% (n=295)
Adult L2 (Turkish)	55% (n=99)	64% (n=102)	25% (n=102)
Adult L2 (M. Arabic)	80% (n=150)	17% (n=159)	17% (n=160)

L2 children and L1 children are generally highly accurate on both VO and OV orders. This contrasts with the L2 adults who perform relatively well on the VO order but poorly on subject-verb inversion; the Moroccan Arabic speakers also perform poorly on the OV condition. Differences between the L2 children and L2 adults (both language groups together) on the latter two conditions are significant ($\chi^2 = 39.58$ and 48.16, df = 1, $p < .001$). For the L2 adults, there is a clear effect of L1 transfer: the Turkish-speaking L2ers are better on the OV order than the Moroccan Arabic/Tarafit-speaking L2ers and vice versa on the VO order. A within-group comparison of the adult data shows that as proficiency increases, L2 subjects become more targetlike. Blom and Polišenská note that the L2 adults nevertheless make more errors than the L2 children and they overuse SVO, that is, they use SVO where the OV order or subject-verb inversion is required.

Blom (2003) (amongst others) has argued that at the point at which L1 children move from a system where finiteness is realised optionally (an Optional or Root Infinitive stage) to a targetlike system with obligatory finiteness marking, they make use of dummy auxiliaries as a means of marking finiteness before they obligatorily raise the lexical verb to the V2 position. An example of such a dummy auxiliary is given in (25).

(25) koe gaat rijden (Matthijs 2;4.24)
cow goes drive
'Cow is driving'

(Blom 2003:196, ex. 116b)

The L2 children and L2 adults tested by Blom and Polišenská were found to differ significantly in the number of dummy auxiliaries used. The results are given in Table 3, for main clause and embedded clause contexts, respectively.

Table 3. Blom and Polišenská (to appear, Table 4): Development of dummy auxiliaries for L1 children, L2 children and L2 adults

Group	% dummy auxiliaries in main clause contexts	% dummy auxiliaries in embedded clause contexts
Child L1 – 3-year-olds (n=10)	23% (n=145)	3% (n=141)
Child L1 – 4/5-year-olds (n=14)	3% (n=210)	0% (n=210)
Child L2 – 6-year-olds (n=11)	57% (n=133)	13% (n=136)
Child L2 – 7-year-olds (n=16)	28% (n=233)	4% (n=221)
Adult L2 (n=11)	1% (n=213)	

Adult L2ers hardly ever use dummy auxiliaries, whereas L2 children do. Furthermore, like L1 children, L2 children generally use dummy auxiliaries in main clause contexts only. More specifically, the older L2 children, the 7-year-olds, use dummy auxiliaries at approximately the same rate as the younger L1 children, the 3-year-olds. Given that these L2 children have been exposed to Dutch for a comparable period (that is, for three years between the ages of four and seven years), Blom and Polišenská suggests that the L1 children and L2 children follow a similar developmental path in that this takes around the same amount of time to reach the dummy auxiliary stage. Note, however, that although the general pattern is the same, the younger (6-year-old) L2 children appear to produce dummy auxiliaries in the OV context considerably more often than the L1 younger (3-year-old) L1 children (compare 13% with 3%).

Finally, Blom and Polišenská examine the types of substitution errors which L2ers make, that is, the forms which subjects overuse when they make errors with finite verbal inflection. The results are given in Table 4, with comparable L1 data taken from de Haan (1996). These rates of substitution were calculated by dividing the number of errors by the number of contexts in which an error could occur. For example, for the 3SG form-*t*, the number of times this form was used erroneously was divided by the total number of contexts for 1SG and all plural forms.

Table 4. Blom and Polišenská (to appear, Table 6): Substitution of verbal suffixes in L2 children and L2 adults

Group	Substitute –*en* (in all SG contexts)	Substitute –*t* (in 1SG, all PL contexts)	Substitute –*ø* (in 2SG, 3SG, all PL contexts)
Child L1	8% (n=180)	74% (n=77)	76% (n=169)
Child L2	33% (n=143)	75% (n=36)	60% (n=151)
Adult L2	48% (n=196)	49% (n=41)	53% (n=198)

When the L2 children substitute one form for another, it is either –*t* or –*ø*. This is similar to the L1 children, who also prefer –*t* or –*ø*. Note, however, that unlike the L1 children, the L2 children also make use of –*en*. The L2 adults do not have any preferred substitute. Thus, Blom and Polišenská argue, the difference between the L2 children and L2 adults boils down to a difference in the use of *–en* as a substitute. The authors note that this finding concurs with previous results (Prévost and White 2000).

To summarise: Blom and Polišenská mostly observe differences between L2 children and L2 adults in their acquisition of the morphosyntactic properties of finiteness. Although the two groups make use of two of the same substitution forms, on most variables the L2 children pattern more like the L1 children than the L2 adults. With respect to transfer, Blom and Polišenská note that the lack of L1 transfer in the child L2 data (cf. L2 adults – see Table 2) may be a consequence of the children being beyond the relevant stage. They claim, however, that there are two key differences between the two L2 groups which cannot be attributed to L1 transfer, namely: (i) the use of dummy auxiliaries by the L2 children but not by the L2 adults, and (ii) the overuse of SVO order by the adults but not by the L2 children. According to Blom and Polišenská, these two observations confirm the critical period hypothesis (cf. §1.2.1.1), that is, for the variables in question, adult L2 acquisition appears to be fundamentally different from child L2 acquisition.

That there are clear differences between the L2 children and L2 adults in the data (as they are currently analysed) is evident. However, before such a far-reaching conclusion can be drawn, more fine-grained analyses must be carried out. Specifically, it is necessary to systematically tease apart the roles of L1 transfer and proficiency and the issues of development and ultimate attainment. For example, in addition to considering the verb placement results for the different source language groups separately, within each of these groups (i.e. Turkish-speaking L2 children, Turkish-speaking L2 adults, etc.), it is also necessary to consider the role of proficiency. This will allow an analysis of the L2ers' development on each of the respective variables. With respect to the proficiency level of the different L2 groups, there is one potential cause for concern and that is the observation that the child L2 group is generally much more proficient than the adult L2 group. As noted above, proficiency was measured using a sentence repetition task and on the basis of the scores obtained, subjects were divided into three proficiency groups. Amongst the L2 children, there were just two subjects in the low proficiency group, 15 in the moderate proficiency group and 14 in the high proficiency group. This contrasts significantly with the relative distribution of the L2 adults across these three groups, *viz.* nine in the low, seven in the moderate and two in the high proficiency groups (Blom 2005). That the L2 adults are generally much worse than the L2 children on the experimental task could in part be explained by this difference. One further complicating issue concerning proficiency is the fact that the proficiency task also incorporates a measure of subjects' knowledge of word order (see §4.4.1.2.1 for details). This means that the experimental tasks and the proficiency task could partially be testing the same thing. Finally, it is of course possible that the group data are masking interesting and potentially important differences between individual subjects or groups of subjects. For example, Blom and Polišenská note that when L2 adults inflect the verb with a non-targetlike suffix, they do not have any preferred substitute form. This may be true for these learners as a group, but in order to further justify this claim, a detailed analysis of the individual data is necessary. By carrying out a series of more fine-grained analyses along the lines suggested here, it will be possible to determine whether the observed similarities and differences are as clear-cut as they at first seem. Until these analyses are available, however, Blom and Polišenská's results should be treated with a certain degree of caution.

1.4.3.5 Dimroth (2005)

In a recent paper, Dimroth (2005) examines the acquisition of finiteness and negation in the L2 acquisition of German by two Russian-speakers, aged 8;7 and 14;2. By using longitudinal data, she is able to track the development of these L2ers in considerable detail. The properties of German which she examines are comparable to those investigated by Blom and Polišenská (to appear) and Haberzettl (1999), that is, verb placement and verb form. Specifically, she considers the position of the verb in relation to negation to be an indicator of morphosyntactic finiteness: non-finite verbs appear to the right of negation, whereas finite verbs appear to the left, as illustrated in (26). In main declarative clauses, finite lexical verbs always raise past negation. This contrasts with English, which uses auxiliaries or *do*-support to express finiteness in negated utterances, as indicated by the gloss to the German example. Note that the negator in Russian always immediately precedes the finite verb.

(26) a. Alison **hat** nicht **gehackt**
Alison has not hoed
'Alison hasn't hoed.'

Reviewing previous studies on the acquisition of finiteness and negation in German by L2 adults whose L1 has preverbal negation (e.g. Becker 2005; Meisel 1997; Parodi 2000), Dimroth observes that (i) as soon as finite non-thematic verbs are used, they appear in a pre-negation position; and (ii) the acquisition of the auxiliary system precedes the acquisition of lexical verbs which are finite in both form and position, that is, lexical verbs which are inflected and which are raised past negation. Dimroth claims that the acquisition of the auxiliary system, and in particular, of non-modal auxiliaries, constitutes a crucial step in the acquisition of the functional category system. Specifically, it enables a morphological agreement relation to be established between the auxiliary and the subject and likewise, a head-complement relation between the auxiliary and the lexical verb. It also allows finiteness to be encoded separately from the lexical context of verbs. In this (and other ways), Dimroth considers the acquisition of the auxiliary system a prerequisite for the productive morphosyntactic marking of finiteness on lexical verbs.

Dimroth examines longitudinal spontaneous production data from two (in her terms) non-adult L2ers to determine whether these observations regarding the development of finiteness and negation in L2 German also hold for younger learners. Her subjects are two Russian-speaking sisters, a child (Natsja) and an adolescent (Dascha), who were 8;7 and 14;2 at first exposure (and hence both actually L2 adults on the definition adopted here). Data collection started just three weeks later and files from the first 24 weeks are included in the analysis.

With respect to the position of finite non-thematic verbs in the emerging L2 system, the same observation made above for L2 adults also holds for both the L2 adolescent and the L2 child, that is, modals, auxiliaries and the copula all appear in finite form in pre-negation position from very early on: pre-negation copula, as in (27)-a, is present in the data from the very first recording for both learners, and the adolescent, Dascha, uses pre-negation modals, as exemplified in (27)-b, six weeks after first exposure and the child, Natsja, after five weeks.[23]

(27)	a.	meine	mutter	**ist**	**nicht**	gut	in latein		(Dascha, week 2)
		my	mother	is	not	good	in latin		
	b.	morgen	**ich**	**kann**	nicht	to	Apostelgymnasium		(Dascha, week 3)
		tomorrow	I	can	not	to	Apostelgymnasium		

(Dimroth 2005: 21-22, ex. 29a, 30f))

We now turn to the relationship between the acquisition of auxiliaries and the acquisition of finite lexical verbs. In the first 13 weeks of data collection, the adolescent, Dascha, produces only one lexical verb in a pre-negation position, namely *haben* 'to have'. This verb appears exclusively with the suppletive form, *kein* 'no/not any', as in (28), and this form of the negator appears with no other lexical verb before auxiliaries are acquired. Given its restricted use, Dimroth argues that this raised lexical verb is unproductive.

(28)	ich	**habe**	**kein**	fotos	meine	vater	hier	(Dascha, week 4)
	I	have	no	photos	my	father	here	

(Dimroth 2005: 23, ex. 32d)

The auxiliary *haben* 'to have', illustrated in (29), first appears in week 14 and it is used productively from week 16, where 'productive' should be understood as it occurring with more than one past participle. To be specific, in week 16, Dascha produces eight auxiliaries, five of which are used with different past participles. This also coincides with the productive use of tense marking; here,

[23] Examples are replicated from the original, i.e. with word-for-word glosses only. Weeks refer to the number of weeks after the first recording.

'productive' should be understood as the occurrence of a past tense suffix with more than one verb. The first non-formulaic lexical verbs raised past negation appear in week 14 (type/token ratio = 4/5). No lexical verbs are produced in this position in weeks 15 and 16, but in week 17, a variety of lexical verbs frequently appear in pre-negation position (type/token ratio = 7/17), as in (30). This is one week after auxiliaries are used productively.

(29)	ich	**habe**	**gesehen**	Mona Lisa	(Dascha, week 16)
	I	have	seen	Mona Lisa	
					(Dimroth 2005: 18, ex. 24a)

(30)	sie	**gucken**	**nicht**	wie	russen	(Dascha, week 17)
	they	look	not	like	Russians	
						(Dimroth 2005: 24, ex. 35f)

Thus, Dimroth claims, Dascha follows the same pattern as the L2 adults in previous studies, only raising lexical verbs across negation once auxiliaries have been acquired.

The data from the younger learner, Natsja, do not follow this pattern. In contrast to her older sister, Natsja produces lexical verbs in pre-negation position in the very first recording (type/token ratio = 2/2) and after a short gap in weeks 3 and 4 with no lexical verbs, this gradually increases (type/token ratio = 2/6 in week 5, 1/2 in week 6, 7/10 in week 7, 2/12 in week 8, etc.). Examples are given in (31).

(31)	a.	heute	ich	**geht**	**nicht**		(Natsja, week 5)
		today	I	go	not		
	b.	sie	**kauft**	**nicht**	torte		(Natsja, week 7)
		she	buys	not	pie		
							(Dimroth 2005: 31-32, ex. 48c, 48l)

As noted above, non-thematic verbs are produced in pre-negation position from the first recording. Auxiliaries only become productive, however, in week 11, when Natsja produces 11 tokens, six of which are used with different past participles. For the child learner, then, the acquisition of auxiliaries does not precede the acquisition of finite lexical verbs. Thus, Dimroth concludes, in their acquisition of finiteness and negation, this child and adolescent L2er pass through different development stages, *contra* Schwartz's (2003) predictions (essentially the same as those in Schwartz (1992) outlined above).

There are, however, several problems with the data, the analysis and the comparability of the two subjects, all of which undermine Dimroth's claims. Recall that Dimroth maintains that in the adolescent L2er's data, lexical verbs are only productively raised across negation (in week 17) once auxiliaries are acquired (between weeks 14 and 16). Part of the evidence for this claim is the absence of lexical verbs raised across negation in weeks 15 and 16. Note, however, that during this period, Dascha also fails to produce lexical verbs which are *not* raised across negation. In other words, she does not produce *any* lexical verbs in combination with negation. It is thus impossible to know whether lexical verbs are raised across negation during this period. Hence, to state that they are only acquired in week 17 is potentially inaccurate as absence of evidence does not constitute evidence of absence. It could be that lexical verbs are not raised across negation until week 17, but because Dascha does not produce any lexical verbs in combination with negation, raised or not, one simply cannot make any claims about the nature of her interlanguage grammar at this point. Missing data in this way is unfortunately one of the disadvantages of using spontaneous production data.

Dimroth's claim regarding the relative order of acquisition of the auxiliary system and lexical verb raising is also undermined by the paucity of data for auxiliaries. Recall that Dimroth claims that for Dascha, auxiliaries are productive by week 16, whereas lexical verbs are raised across negation productively one week later. The actual difference between the number of auxiliaries and the number of lexical verbs raised across negation before week 17 is small enough to at least wonder about its significance. Before week 17, Dascha produces 11 auxiliaries of which six are different and five lexical verbs raised across negation of which four are different. Coupled with the lack of relevant tokens for lexical verbs in weeks 15 and 16, this somewhat weakens Dimroth's claim that for Dascha, the adolescent learner, auxiliaries are acquired before lexical verb raising.

One further problematic aspect of Dimroth's analysis is that in addition to differing in terms of their age, her two subjects also differ with respect to previous linguistic knowledge: for the adolescent L2er, German is her second L2 because she already knows English (at what level is not stated, but example (27)-b demonstrates use of this L2 in her German). This means that although on several variables, such as length of exposure, these two sisters will be truly comparable, they differ in two crucial aspects, namely age and knowledge of another L2. Therefore, any observed differences between the two subjects in their linguistic development could potentially be due to either of these two variables. More specifically, given that English makes use of auxiliaries to express finiteness in negated utterances, Dascha's knowledge of English could potentially contribute to her earlier acquisition of the auxiliary system in German.

Thus, although the longitudinal nature of Dimroth's study should be applauded, the fact that the older L2er already knows another L2, coupled with the absence of relevant data at crucial points, means that Dimroth's strong claim that the two L2ers go through different developmental stages cannot be maintained.

1.4.3.6 Summary

Developmental sequences were one of the first types of data to be investigated as part of a child L2 ~ adult L2 comparison and they continue to be the focus of current research. In recent research, there is evidence to suggest both that these two groups of learners progress through the same stages (Gilkerson 2004) and that they are different (Blom and Polišenská to appear; Weerman et al. 2003), although methodological weaknesses in some of these studies may in part be responsible for these findings. All of the studies reviewed are based on production data, elicited or spontaneous. None of them investigate what L2 children or adults consider to be ungrammatical. In fact, as the summary table in Appendix A shows, there are relatively few child L2 studies – just 12 out of 49 – which examine types of data other than elicited or spontaneous production. Of these 12, there are only four which employ a grammaticality judgement task and none uses a truth value judgement task. The present study, which, in addition to elicited production, also uses two truth value judgement tasks, attempts to fill (at least some of) this gap.

The review in this section has also highlighted some of the methodological issues which arise when child L2 acquisition is compared with adult L2 acquisition. Regarding the comparability of these two groups, we have seen that it is necessary to ensure that – in so far as feasible – these two groups are as comparable to each other on as many variables as possible, if a child L2 ~ adult L2 comparison is to be valid. In particular, it is essential to have some means of knowing which child(ren) to compare with which adult(s). This is not always clear in some studies. How this issue is addressed in the present study is discussed in Chapter 4.

1.5 Child L2 ~ Child L1 : Differences and similarities

This section reviews studies which specifically compare child L2 data with data from L1 children (either collected as part of the study in question or from the literature). It addresses the same three topics as in the preceding section on child L2 ~ adult L2 comparisons: age, L1 transfer (and the initial state) and developmental sequences.

1.5.1 Age (at time of testing)

The extent to which L1 and L2 children differ in terms of their age (at time of testing) partly depends on the property of language in question (cf. discussion in §1.4.1). Put simply, the L2 child will either be younger or older than the age at which the relevant property of language is acquired by the L1 child. The prediction is that if a particular developmental stage in L1 acquisition, occurring at age X, is maturationally determined, then any L2 child who is older than age X should not pass through that developmental stage. Given that most aspects of morphosyntax and phonology are in place by age four and that child L2 acquisition is defined as L2 acquisition where first exposure is between the ages of four and seven, it will generally be the case that L2 children differ in a non-trivial way from L1 children, namely they will already have linguistic knowledge which they could transfer to their interlanguage grammar.

The implications of maturational age depend on the importance attributed to it in a given acquisitional theory. As noted above (§1.2.2), there exist acquisitional accounts which attribute particular developmental stages in L1 acquisition to maturation. Such accounts entail very specific predictions about whether L2 children (and adults) will pass through similar stages as L1 children. Of course, if maturation, i.e. the particular age X, is not considered a crucial factor in the developmental stage in question, it does not matter whether L2 children pass through the same developmental stage, albeit somewhat later in chronological age, as L1 children. Studies which have examined child L2 acquisition in a maturational light include Henry and Tangey (1996), Ionin and Wexler (2002), Park (2000) and Schwartz (2004), amongst others; some of these are discussed below in §1.5.3.

What about properties of language which are acquired after the age of four years? Here, the logic is essentially the same: any L2 child who is younger than the age at which a maturationally determined developmental stage purportedly occurs in L1 acquisition is predicted to not yet be in that stage. Thus, for example, if a particular property of the target language is only acquired by L1 children at say, age eight, L2 children will not necessarily be expected to acquire that property before age eight Note that this logic in fact holds for any late-acquired target language property without it necessarily having to be maturationally determined. Such delayed acquisition might, for example, relate to limited input, pragmatic and/or lexical development.

To summarise: Age is a crucial variable in the child L2 ~ child L1 comparison. Whether these two groups are to be considered similar or different in terms of age depends on the property of language in question and the age at which it is acquired by L1 children. In virtually all previous studies, the relevant property of the target language is acquired before age four, and hence L2 children are considered to differ from L1 children. In other cases, L1 children remain non-targetlike until a later age and therefore, L2 children younger than this age at time of testing can be considered comparable to L1 children of the same age. This is the type of situation which will be observed in Chapter 6.

1.5.2 L1 transfer and the initial state

L2 children are clearly different from L1 children because they already have knowledge of another language and, as we have seen in §1.4.2, this has been found to manifest itself in the interlanguage grammar as L1 transfer. Most of the evidence that L2 children exhibit transfer was reviewed in that section. Here, a number of studies which address this issue in terms of an explicit comparison of L1 and L2 children will be presented. First, Gavruseva (1998) investigates whether the types of errors L2 children make in their acquisition of long-distance *wh*-questions is the same as those of L1 children or whether they result from L1 transfer. Second, we (re)turn to the issue of functional categories (cf. §1.4.2) in a discussion of Grondin and White's (1996) study of child L2 French.

1.5.2.1 Gavruseva (1998)

Gavruseva examines the acquisition of (matrix and long-distance) questions with *wh*-possessive phrases, such as:

(32) a. **Whose spade** did you borrow?
b. ***Whose** did you borrow **spade**?
c. ***Who** did you borrow**'s spade**?

In English matrix and long-distance questions, the noun in a *wh*-phrase is obligatorily pied-piped with the *wh*-word, as illustrated for the matrix question in (32)-a. Extraction of *whose* or *who* on its own, as in (32)-b and (32)-c, respectively, is ungrammatical. It has been observed, however, that L1 children pass through a developmental stage where they allow *wh*-extraction of this type. This is restricted to long-distance questions and *whose*-questions only. Gavruseva reports on data from L1 children (aged 4;5-6;0) who, in an elicited production task, produced sentences such as:

(33) a. **Who** do you think**'s cat** came up on the building?
(Whose cat do you think came up on the building?)
b. **Who** do you think**'s food** the baboon tried?
(Whose food do you think the baboon tried?)
(Gavruseva 1998:236, ex. 235)

Gavruseva asks whether L2 children pass through a similar developmental stage as L1 children and whether L1 transfer plays a role in the child L2 acquisition process. Using an elicited production task, she elicited a variety of questions from two Russian-speaking children acquiring English. The first child, Nadia, exposed to English at age 4;9, was age 6;5 at the time of testing. The second child, Alex, was 5;11 at time of testing and 5;2 at first exposure. Russian, their L1, allows adjectival *wh*-possessors to be extracted, as illustrated in (34), but this is subject to certain constraints. Firstly, the *whose* in *whose*-N phrases can never be long-distance extracted from tensed embedded clauses, but it can be extracted (i.e. split) in matrix clauses. Secondly, extraction is possible with other types of *wh*-phrase, such as *which N, how many N*. Note, however, that Russian does not have a direct equivalent of questions of degree, such as *how big*. Rather, questions of degree are realised by a *which N* phrase marked for genitive case.

(34) a. **Chju** **shapku** on poteryal
whose-fem.sg.ACC hat-fem.sg. ACC he lost
'Whose hat did he lose?'

b. **Chju** on poteryal **shapku**
whose-fem.sg.ACC he lost hat-fem.sg. ACC
'Whose hat did he lose?

(Gavruseva 1998:235, ex. 232)

Data were collected from the two children over a number of sessions. The distribution of their responses is given in Table 5.

Table 5. Gavruseva (1998): Distribution of matrix and long-distance *whose*-questions

Subjects	Matrix		Long-distance		Other[24]
	Target	Split	Target	Split	
Nadia	1	6	-	24	11
Alex	0	0	-	4	0

As Table 5 shows, Alex produced very few questions, just four in total. Nadia, on the other hand, produced many more, both split and targetlike forms. The types of *wh*-phrase which were extracted are as follows. First, *wh*-words were extracted out of main and embedded clauses, as shown in (35) and (36), respectively. This is consistent with transfer from Russian in the sense that in this language questions can be split in both clause types. However, in Russian, *whose* cannot be extracted from tensed embedded clauses (cf. (36)). Second, like L1 English children, Nadia consistently pied-piped the noun in *which N* and *how many N* phrases. Unlike L1 children, however, she split *wh*-phrases with questions of degree, as in (37). Gavruseva claims that this cannot be due to L1 transfer because Russian does not have questions of degree. Furthermore, *wh*-extraction is optional in Russian and this is not reflected in Nadia's L2 production: in the first three sessions, all questions of degree are split. In the fourth, Nadia produces one split question of degree and three targetlike (pied-piped) questions and in the fifth session, all *wh*-questions, except for *whose N* questions, are targetlike.

(35) a. **Who** did you not like **someone's drink**? (Nadia 6;5)
b. **Who** did you like **dessert**?

(Gavruseva 1998:242, ex. 218)

(36) a. **Who** do you think Pocahontas likes **the chair**? (Nadia 6;5)
(Whose chair do you think Pocahontas likes?)
b. **Who** do you think the cheetah tried **drink**? (Alex 5;11)
(Whose drink do you think the cheetah tried?)

(Gavruseva 1998:241, ex. 217)

[24] It is not clear what type of utterances are included in the 'other' category.

(37) Experimenter: I was wondering how long his neck really is. (Nadia 6;5)
Can you find that out for me?
Nadia: **How** your neck is **long**?
(Gavruseva 1998:243, ex. 219)

Thus, Gavruseva claims, in these child L2 data, there is some evidence for L1 transfer, that is, as in Russian, *whose N* phrases are split in matrix questions. This is not found in the child L1 data. There is also evidence that the L2 children behave like L1 children, as they do not extract from *which N* and *how many N* phrases. There are also responses which neither directly result from L1 transfer nor are they identical to the child L1 data. For example, in addition to splitting *whose N* questions, which are the only type of question to be split by L1 children, Nadia also splits questions of degree. The Russian equivalent of questions of degree involves a *which N* phrase, but this cannot be extracted. Although the different properties of *wh*-extraction in Russian allow L1 transfer to be differentiated from a developmental stage comparable to L1 acquisition in an interesting and sophisticated fashion, the fact that both L1 transfer and regular development allow *wh*-extraction mean that either one could be a source of (some of) the overgeneralisations which are observed in (at least one of) these children. One further piece of important information which is missing from this paper is whether these children have acquired the constraints of *wh*-extraction in Russian, i.e. whether there is actually something to transfer. It should be noted, however, that Gavruseva's (1998) study is not unusual in this regard: most child L2 studies (the present included) examine the children's development in their L2 only (but see, for example, Aarssen (1996), Boeschouten and Verhoeven (1986), Bos (1997) and Jin (2003) who document their child L2 subjects' L1 development as well).

To summarise: Gavruseva's study investigates the role of L1 transfer and the question of whether L2 children pass through the same developmental stages as in L1 acquisition. To the extent that it is possible to tease these two issues apart, there was evidence for both. This study also highlighted the important point that in studying L2 children's acquisition of (quite complex) properties of the target language, it is important to make sure that, if claims about the role of L1 transfer are to be made, the equivalent property in the children's L1 has indeed been acquired.

1.5.2.2 Haznedar (2001; 2003)

The issue of transfer forms a crucial part of the debate about the initial state in L2 acquisition. This debate revolves around the question of which (functional and/or lexical) projections are present in interlanguage grammars at the onset of acquisition. Some researchers claim that only lexical categories are present (e.g. Vainikka and Young-Scholten 1994; 1996a; 1996b), whereas others maintain that the full repertoire of functional categories is available, transferred from the L1 (Schwartz and Sprouse 1994; 1996). Like many issues in L2 acquisition, the origins of this debate can be found in L1 acquisition research. There, it has been claimed (e.g. Clahsen, Eisenbeiss and Penke 1996; Vainikka 1993/1994) that young children start by projecting lexical categories only (plus possibly one underspecified functional category – (Clahsen 1990/1991)), building up their inventory of functional categories, on the basis of the input which they hear. This contrasts with the *Full Competence* approach (e.g. Poeppel and Wexler 1993), which entails that all functional categories are present from the onset. Child L2 data are relevant to this issue in L1 acquisition because if L2 children were observed to pass through a stage where they produced utterances consistent with the presence of lexical categories only, this (according to some researchers at least) would provide evidence for a weak continuity approach and against a strong continuity approach to acquisition. Of course, the whole issue of the presence or absence of functional categories in early grammar depends on the extent to which the absence of, for example, inflectional morphology can be assumed to reflect the absence of the functional category which hosts this morphology. If such inflectional morphology and other such instantiations of functional categories are present in early child L2 utterances, however, they constitute clear evidence against a structure-building approach (to child L2 acquisition at least).

The debate surrounding the initial state is one area where child L2 data have been put to great use. In two recent articles, Haznedar (2001; 2003) has convincingly demonstrated that in the L2 English data of her Turkish-speaking subject (see §1.4.2), there is evidence for both the IP and CP projections from a very early stage. With respect to the IP, for example, she observes that copula *be* and auxiliary *be* appear very early on (after just one month of exposure) and they are used in the correct syntactic context. This contrasts with 3SG –*s*, which appears much later (after six months of exposure) and is regularly omitted. This omission pattern persists for a relatively long period of time, taking another ten months of exposure before -*s* is supplied in 70% of obligatory contexts. This contrast, Haznedar (2001) argues, goes against the structure-building approach of Vainikka and Young-Scholten (1994; 1996a; 1996b), which assumes that missing functional morphology reflects missing functional categories. The initial absence of 3SG –*s* would, on their account, indicate that the functional category which hosts inflectional morphology, IP, were missing. This cannot be the case, however, as other IP-related elements, namely copula and auxiliary *be*, are already used productively and accurately.

Haznedar (2003) presents evidence that the CP projection is also available from early on. She shows that in Erdem's data, yes-no questions, *wh*-questions and *wh*-complement clauses are generally correct from their earliest occurrences and that embedded clauses with the complementisers *because* and *if* and with the verb *want* also appear relatively early. Furthermore, *contra* the structure-building approach, CP-elements are acquired before the IP elements of tense and agreement morphology.

1.5.2.3 Grondin and White (1996)

Grondin and White's study also addresses the issue of whether functional categories are available in early child L2 acquisition. They examine longitudinal production data from two English-speaking boys acquiring French. Their first exposure to French was in a French-speaking bilingual nursery at age 4;9 for the first boy, Kenny, and at 4;5 for the second boy, Greg. During this first year, they produced very few spontaneous utterances in French, however. They subsequently started attending a French-speaking kindergarten, at age 5;10 and 5;6, respectively. This is when data collection began. (Note that the number of months' exposure given in the examples is measured from this point.)

Grondin & White find evidence for the presence of the functional categories, DP and IP, in the earliest stages of the available data. With respect to the DP, they observe that from the earliest recordings, the children consistently produce nominals including a variety of determiners, as illustrated in (38). Greg produces determiners in obligatory contexts at a rate ranging from 86% (24/28) in the first recording (five months' exposure) to 98% (165/169) in the final recording; for Kenny, the equivalent figures are 67% (4/6) and 96% (116/121), respectively. Furthermore, the children's use of the dummy case-marking preposition *de* 'of' in early recordings, exemplified in (39), suggests that they are aware of the requirement that every DP receive case (the *Case Filter*).

(38) a. **Le** lion mange **les** giraffes (Greg, 5 months' exposure)
the lion eats the giraffes
'The lion eats the giraffes.'

b. Ça c'est **un** giraffe
that it's a giraffe
'That's a giraffe.'

c. Ça c'est **la** maman
that it's the mother
'That's the mother.'

(Grondin and White 1996:7, ex. 2a-c)

(39) a. fête **de** Halloween (Kenny, 2 months' exposure)
party of Halloween
'Halloween party.'

b. Ça c'est **de** toi
that it's of you
'Is that yours?'

(Grondin and White 1996:9, ex. 5c-d)

Grondin and White also find evidence for the presence of IP. The data attest both morphological and syntactic reflexes of this functional category. From the earliest recordings, both children produce a variety of verb forms with different verbs (both lexical and auxiliary). Like in L1 French acquisition, very few person-agreement errors are made and person agreement was acquired before number agreement. Grondin and White argue that the children's use of subject clitics, which they assume to realise subject-verb agreement and therefore to be situated in Infl, also demonstrates knowledge of IP: subject clitics are produced from the first recording for each subject; they are almost always used with finite verbs, as the target language requires; they always show targetlike person agreement (as evidenced by subject-clitic doubling and the context); and finally, they are never used in a non-targetlike (non-subject) position.

The most convincing piece of syntactic evidence Grondin and White cite for the presence of IP is the targetlike placement of verbs with respect to negation. Assuming that the negator *pas* 'not' occurs between VP and IP, any verbs which occur to its left are assumed to occupy Infl. Finite verbs produced to the left of negation are therefore evidence that the IP projection is present in a child's grammar. For both Kenny and Greg, there is a highly significant association between verb placement and finiteness: finite verbs are correctly placed to the left of negation (at a rate of 98% (449/457) for Greg and 97% (321/331) for Kenny) and non-finite verbs are placed to the right (at a rate of 94% (15/16) for Greg and 95% (21/22) for Kenny). Correct verb placement with respect to negation is observed from the earliest available recordings. Examples are given for finite verbs in (40) and non-finite verb in (41).

(40)	a.	Ça	c'**est**	**pas**	le	ferme		(Greg, 5 months' exposure)
		that	it's	not	the	farm		
		'That's not the farm.'						
	b.	Non	j'**ai**	**pas**	joué	avec		(Greg, 9 months' exposure)
		no	I've	not	played	with		
		'No, I didn't play with.'						
	c.	Le	giraffe	**peut**	**pas**			(Kenny, 5 months' exposure)
		the	giraffe	can	not			
		'The giraffe can't.'						

(Grondin and White 1996 18, ex. 14a,b,d)

(41)	a.	Moi	je	**pas**	**jouer**	avec	le auto	(Greg, 11 months' exposure)
		me	I	not	play	with	the car	
		'I don't play with the car.'						
	b.	Non	**pas**	**jouer**				(Kenny, 5 months' exposure)
		no	not	play				
		'No, I don't want to play.'						
	c.	**Pas**	**ouvrir**	ça				(Kenny, 5 months' exposure)
		not	open	that				
		'Don't open that.'						

(Grondin and White 1996 18-9, ex. 15)

Grondin and White thus conclude that in their acquisition of French, these two L2 children have DP and IP from the earliest stages at which they were observed (see also, for example, Gavruseva (2000) and Henry and Tangey (1996) for similar conclusions).[25] With respect to transfer, Grondin and White suggest that it is found where the L1 and L2 are similar, for example, in the use of determiners or the formation of *wh*-questions, but for those areas of the target language grammar for which there are no L1 equivalents, such as verb raising and the use of clitics, child L2 acquisition proceeds along the same lines as child L1 acquisition.

1.5.2.4 Summary

The data which have been reviewed in this section, in combination with those discussed in §1.4.2, show that child L2 acquisition involves L1 transfer and consequently is different from child L1 acquisition. With respect to the initial state, the data examined here suggest that at least some functional categories are present from the earliest stages of child L2 acquisition. The extent to which this can be considered to be different from or similar to child L1 acquisition of course depends on the position one takes in the weak vs. strong continuity debate.

1.5.3 Developmental sequences

In this final child L2 ~ child L1 section, we return to the topic of developmental sequences. As mentioned in §1.4.3, probably the most well-known developmental sequence data comparing different groups of learners are the grammatical morpheme studies. These have been discussed in some detail in the aforementioned section, where they were shown to be problematic in several respects; hence, they will not be discussed any further. Rather, the discussion here focuses on one area which has been a particularly fruitful area of research in child L2 acquisition, namely Optional or Root Infinitives.

1.5.3.1 Optional/Root Infinitives

L1 children (both monolingual and bilingual) have been observed to pass through a stage where their declarative main clauses regularly contain non-finite forms when a finite verb is required, so-called Optional or Root Infinitives (OIs/RIs). Examples are provided in (42).

---

[25] The evidence for the presence of CP in these children's productions is ambiguous and hence no conclusions can be drawn about the availability of this functional category.

(42)	a.	die	helemaal	kapot	maken	(Dutch, Niek 3;1.16)
		that	altogether	broken	make$_{INF}$	
						(Wijnen 1997)
	b.	dormir	là	Michel		(French, P 2;2.1)
		sleep$_{INF}$	there	Michel		
						(Deprez and Pierce 1994)
	c.	ein	stift	haben		(German, Annie 2;3.12)
		a	pen	have$_{INF}$		
						(Unsworth 2003b)
	d.	Annie	help			(English, Annie 2;6.21)
						(Unsworth 2003b)

It has been found that in child L1 acquisition, these non-finite forms exhibit certain properties: firstly, in some languages, they co-occur with null subjects and any overt subjects generally show default case-marking (see e.g. Haegeman 1995; Krämer 1993; Phillips 1995); secondly, they are restricted to non-finite positions (e.g. Poeppel and Wexler 1993); furthermore, despite the fact that non-finite forms are used in place of finite forms, when tense/agreement morphology is used, it is largely accurate (e.g. Poeppel and Wexler 1993). Two of the accounts put forward to explain this phenomenon, Rizzi's (1993/1994) *Truncation Hypothesis* and Wexler's (1998) *Very Early Parameter Setting* account, are both couched in terms of maturation (the precise details of which are not important for the present discussion). Both propose that the OI/RI stage in (L1) child language is maturationally driven and that by about the age of three, children leave this developmental stage and, consequently, finite verb forms replace the non-finite forms. Such approaches make very clear-cut predictions for child (and adult) L2 acquisition: the L2 child, by definition more mature than the L1 child, should not pass through an OI/RI stage of the same type as the L1 child.

In order to test this prediction, studies have sought to establish first of all, whether L2 children use non-finite forms in place of finite forms, and secondly, whether the morphosyntactic contingencies listed above also hold. It has been claimed that only if both of these hold can L2 children be considered to pass through an OI/RI stage (Ionin and Wexler 2002; Schwartz and Sprouse 2002). We will discuss the latter three properties in reverse order, but first, we consider whether L2 children use non-finite forms instead of finite forms. There is plenty of evidence to show that this is indeed the case (Armon-Lotem 1998; Haznedar and Schwartz 1997; Ionin and Wexler 2002; Prévost 1997; 2003; Schwartz and Sprouse 2002; Tran 2005a; 2005b).[26] (43) exemplifies non-

[26] Henry and Tangey (1996) claim that there is no evidence for an OI/RI stage in their data. This claim is impossible to verify, however, given that they only provide examples and no quantitative data. Furthermore, they provide no information about how long their subjects had

finite forms in the L2 German of an English-speaking child, aged 3;2 to 3;7 at the time of testing, taken from Prévost (2003), and (44) gives examples of a variety of non-finite forms produced by the Russian-speaking children acquiring English as L2 in Ionin and Wexler (2002).[27, 28]

(43) a. kaputt **machen** (Cindy, sample 3, day 7)[29]
broken make$_{\text{INF}}$

b. **sitzen** da (Cindy, sample 10, day 33)
sit$_{\text{INF}}$ there

c. nicht **kemacht** (=gemacht) (Cindy, sample 9, day 30)
not done

(Prévost 2003:78, ex.79-11)

(44) a. Third person *–s*
girl **play** with toy (DA, age 9;7)

b. Past tense *–ed*
one time I **watch** this movie (AY, age 10;4)

c. *Be* auxiliary
here she **ø making** a cake (AT, age 6;5)

d. *Be* copula[30]
Mary **ø** so funny (OL, age 6;10)
(Ionin and Wexler 2002:106, ex.102)

Regarding the final morphosyntactic contingency listed above, there is some evidence that when tense/agreement morphology is used, it is, on the whole, used accurately by L2 children: Haznedar and Schwartz (1997) note an error

been exposed the L2, Irish, and hence it is difficult to assess whether these children could be beyond a possible OI/RI stage.

[27] Note that Prévost's subject would not, strictly speaking, be classed as a child L2er according to the definition adopted here as she was exposed to the TL before the age of four. The reason for setting the lower bound at this age is to ensure that (most of) the L1 is place before L2 acquisition starts. Prévost (2003:77) notes that this is the case for Cindy, however: in her L1 English, she makes virtually no errors in inflectional morphology, the use of auxiliaries and question and negative formation.

[28] The age at first exposure for Ionin and Wexler's subjects in (44) are as follows: DA = 9;1, AY = 9;11, AT = 5;6 and OL = 6;5. Only the latter two subjects would be counted as L2 children on the definitions adopted here. Of the remaining 16 subjects in this study, six would be classed as L2 adults (age of first exposure = 8;7 to 13;8) and two as bilingual L1 children (age of first exposure = 0;3).

[29] Days give the number of days after the first recording. The first recording took place after two weeks' of exposure.

[30] It is not clear how the utterance with a missing copula (as given here) should be differentiated from the utterance 'Mary's so funny'. For this reason, it is not a very good example. It is, however, the only example of a missing copula which Ionin and Wexler provide.

rate of just 2.8% (12/433) for 3SG *–s* in the L2 English of their Turkish-speaking subject, Erdem, over a period of almost 1.5 years. Ionin and Wexler (2002) record similarly low error rates in their L1 Russian/L2 English subjects: 5% for 3SG *–s,* 7% with auxiliary *be* and 5% with copula *be*. Examples of each are provided in (45):

(45) a. Third person *–s*
I **likes** costumes for Halloween for Batman (KI, age 6;10)

b. *Be* auxiliary
the three ducks **is** going (GU, age 3;9)

c. *Be* copula
the two kittens **is** big (MA, age 7;5)
(Ionin and Wexler 2002:105-106, ex.103)

Grondin and White (1996) also note that their two L2 French subjects, Kenny and Greg, produce very few person-agreement errors. Tran (2005a; 2005b) investigated the (instructed) L2 acquisition of German by English-speaking children. She also reports that the children in her study make virtually no errors in tense/agreement morphology (0.6% (1/169) across subjects). Note, however, that the data in this case are rather restricted because the elicitation tasks which she designed elicited 1SG only. All these results contrast significantly with those of Blom & Polišenská (to appear, see § 1.4.3.4 for details) who observed much higher error rates in child L2 Dutch. Across the 27 children (L1: Turkish or Moroccan Arabic and/or Tarafit), when tense/agreement morphology was used, it was used in a non-targetlike fashion in 18% of 847 occurrences. This contrasts with both L2 adults and L1 3-year-olds, who had error rates of 51% (n=272) and 29% (n=145), respectively.

There are two comments to be made in connection with these results. First, as noted above, Tran's experiment elicited 1SG only. The fact that Blom and Polišenská's tasks elicited all possible verb forms might explain the discrepancy between these two sets of results: in Blom and Polišenská's experiment, subjects had to produce a greater variety of verb forms which meant that there were more opportunities to commit errors. Second, the restricted nature of tense/agreement morphology in English might also account for the lower error rate in Haznedar and Schwartz (1997) and Ionin and Wexler (2002) (see Blom (2003) for similar discussion regarding the restricted nature of English tense/agreement morphology having consequences for cross-linguistic comparisons on OIs/RIs). The evidence is thus mixed regarding L2 children's accuracy on tense and agreement morphology. Most studies show that when it is used, tense/agreement morphology is used in a targetlike fashion, but there is some evidence that tense/agreement errors are committed by some child L2ers. More data are needed from languages with a range of tense/agreement marking before a firmer conclusion can be drawn on this matter, however.

Whether the second of the morphosyntactic contingencies listed above holds, that is, whether non-finite forms are restricted to non-finite positions, is not entirely clear in the data currently available. Although several studies report that the same verb form-placement contingency as observed in the child L1 data is also found in the child L2 data (Armon-Lotem 1998; Grondin and White 1996; Ionin and Wexler 2002; Prévost 2003), there is some evidence that L2 children do not always produce non-finite forms in non-finite position. Tran (2005a; 2005b) reports that non-finite forms are used in (finite) V2 position in her data (cf. Prévost 2003, who examines the same language combination of L1 English/L2 German).

The data regarding the co-occurrence of non-finite forms with null subjects (in some languages at least) are also rather ambiguous. Several studies have observed that null subjects do predominantly occur with non-finite forms (Prévost 1997; 2003). For example, Prévost (2003) observes that in the data from his L1 English/L2 German subject, 61.7% (63/102) of non-finite verb forms have a null subject, whereas this is the case for just 8.6% (49/570) of finite verb forms.[31] Examples of each are given in (46) and (47), respectively.

(46)	a.	kaputt	**machen**			(Cindy, sample 3, day 7)
		broken	make$_{\text{INF}}$			
	b.	da	**essen**			(Cindy, sample 18, day 77)
		there	eat$_{\text{INF}}$			
	c.	so	ein	Haus	**bauen**	(Cindy, sample 8, day 26)
		this-way	a	house	build$_{\text{INF}}$	
						(Prévost 2003:84, ex.21)

(47)	a.	**du**	**hast**	gemacht		(Cindy, sample 8, day 26)
		there	have	done		
	b.	(**ich**)	**hab**	keine	gelp	(Cindy, sample 21, day 89)
		(I)	have	no	yellow	
	c.	(**du)**	**muss**	heim	gehen	(Cindy, sample 21, day 89)
		(you)	must	home	go$_{\text{INF}}$	
						(Prévost 2003:84, ex.22)

Haznedar and Schwartz found, however, that although like L1 children, their Turkish-speaking subject produces null subjects in his L2 English, these are

[31] Null subjects appear as soon as verbs are produced. Prévost does not discuss whether the use of null subjects follows any developmental pattern; he states that they occur in almost all recordings from this point on.

restricted to non-finite forms only. This contrasts with child L1 English, where null subjects occur with both non-finite and finite forms. Moreover, unlike what has been observed for L1 children acquiring languages other than English, there is no developmental correlation between the use of null subjects and the production of verbal inflection: by the time Erdem no longer uses null subjects, his verbal inflection is still largely non-targetlike (at this stage four out of 22 (18%) verb forms are inflected). Other studies have found very few null subjects (e.g. Ionin and Wexler 2002) or no evidence for the contingency between null subjects and non-finite forms (Schwartz and Sprouse 2002 on L1 Italian/L2 German). Furthermore, unlike in L1 English acquisition, where subjects in OIs/RIs tend to have default case-marking, the child L2 data report no such contingency. In other words, in child L2 acquisition, subjects are not necessarily more frequently marked for default case; they sometimes also occur with non-default case-marking. (Gavruseva 2000; Haznedar and Schwartz 1997; Prévost 1997; Schwartz and Sprouse 2002).

To summarise: All in all, the data regarding the morphosyntactic contingencies expected if a child L2 RI stage is of the same type as its L1 equivalent are indeterminate. Despite this being one of the most fruitful areas of research in child L2 acquisition, the results are still relatively incongruous. This could be a result of the different types of data which have been used (elicited production vs. spontaneous; cross-sectional vs. longitudinal). It is also possible that different criteria for determining errors could have played a role. Furthermore, the age of first exposure for the subjects in these studies spans a relatively large age range (cf. fns. 27 and 28). By narrowing down this age range, for example, by only examining children exposed to the L2 between the ages of four and seven, or even further, the variability in the results could perhaps be reduced.

1.5.3.3 Summary

The evidence regarding the question of whether L2 children pass through the same developmental stages as L1 children is rather inconclusive. Although it is clear that as a result of L1 transfer, these two groups differ from each other and as such their developmental sequences will differ, the question of whether certain developmental stages found in child L1 development also characterise child L2 development remains largely unanswered. As yet, the child L2 data are rather limited, in terms both of the linguistic phenomena which have been systematically investigated and the language combinations of the L2 children who have been studied. Put simply, more data are needed. This thesis goes some way to meeting this need.

1.6 Desiderata for a child L2 ~ adult L2~child L1 comparative study

I close this chapter by, on the basis of the foregoing review, presenting an annotated list of desiderata for three-way child L2 ~ adult L2~child L1 comparisons of the type which will be carried out in this thesis. Note, however, that most of the points mainly concern the child L2 ~ adult L2 comparison only.

A. HOLD THE L1 CONSTANT

In order to be able to rule out L1 transfer as a source of disparity between the L2 children and the L2 adults, it is essential that these two groups have the same L1.

B. USE SAME METHODOLOGY WITH EACH GROUP

In order to achieve maximum comparability between children and adults, it is necessary to use the same data collection methodology with all groups. Although finding a particular (experimental) method which is suitable for use with both children and adults is admittedly complicated by the cognitive differences between these two groups, such methods do exist. The following tasks, for example, have been used successfully with both child and adult subjects: act-out, elicited production and grammaticality judgement tasks.

C. USE AN INDEPENDENT MEASURE TO FACILITATE BETWEEN- AND WITHIN-GROUPS COMPARISONS

It has been repeatedly stressed in this chapter that in order to ensure that a particular group of L2 children is truly comparable to a particular group of L2 adults, some independent measure of proficiency (or some other reliable variable common to both groups) is required. Such a measure will also allow within-group comparisons to be made, for example, so that developmental sequences can be inferred.

D. CONSIDER L2 CHILDREN'S AGE AT TIME OF TESTING

If the topic of investigation is development, L2 children need to be children at the time of testing, whereas if ultimate attainment is the locus of interest, they should be adults (or at least adolescents), that is, they should have had enough exposure for them to be at their end state.

E. TEST SOMETHING WHICH IS NOT SUBJECT TO INSTRUCTION IN THE L2 CLASSROOM

This point is valid only for comparisons which include L2ers who have been instructed in some way. Testing L2ers' knowledge of an aspect of language which they are instructed upon in the L2 classroom may mean that children and adults' responses in experimental tasks stem from different sources. It is often – though by no means always – the case that within a group of L2 children and L2 adults, it is the older L2ers (adults and possibly older children) who are more likely to have been instructed than the younger L2ers. For example, the L2 children may be immersed in target language schools where they receive no specific L2 instruction, whereas the L2 adults follow language courses especially designed for L2ers. Consequently, the behaviour of the older L2ers could be driven by metalinguistic knowledge rather than their underlying interlanguage grammar, whereas only the latter is likely to be used by younger L2ers.[32]

F. EXAMINE NEW (OR UNDERRESEARCHED) L1/TL COMBINATIONS AND TL PROPERTIES

Several of the studies reviewed in this chapter make use of the same data sets. Although detailed investigations of the same subjects can provide interesting information about the status of various properties of the interlanguage grammar at one and the same time, they potentially lack generalisability, both across L2 children with the same L1/TL combination and across L2 children with other L1/TL combinations. Similarly, although examining the same phenomenon (such as RIs) across different language combinations is clearly informative in terms of crosslinguistic generalisability, in order to obtain a more complete picture of how child L2 acquisition proceeds, more data are needed on different topics. Specifically, given the predominantly morphosyntactic focus of previous work, it would be desirable to address, for example, properties of language at the interface between syntax and discourse or between syntax and semantics. This thesis goes some way to redressing this gap.

---

[32] Note that this is not to say that L2 children and L2 adults who have not been instructed do not differ in terms of metalinguistic knowledge: it is generally the case that metalinguistic knowledge develops with age. The point is that targeting a linguistic property which is subject to instruction is more likely to encourage the use of such knowledge amongst subjects who have been instructed. Focussing on subtle properties of the target language which are underdetermined by the input – the tactic which is taken here – should go some way toward counteracting the more general difference between younger and older subjects in terms of their metalinguistic knowledge.

1.7 Summary

This chapter started by outlining the motivation for studying child L2 acquisition. It detailed how cross-group comparisons with adult L2 acquisition and child L1 acquisition can inform research on both the native and non-native acquisition process. A detailed overview of recent child L2 literature was presented, focussing on the issues of age, L1 transfer and developmental sequences. It was shown that there are reasons for L2 children to be considered both similar to and different from both L2 adults and L1 children.

With respect to chronological age, the property of the target language in question and when it is acquired by L1 children were seen to be crucial in this regard. One of the clearest results of recent research is that like adult L2 acquisition, the initial state of child L2 acquisition is also characterised by L1 transfer. Furthermore, the available evidence suggests that most functional projections are present from the earliest stages of development. The data regarding developmental sequences are less clear. Similarities and differences have been attested between both L2 children and L2 adults and between L2 children and L1 children. It was noted that the available data are rather scarce. Furthermore, they involve many of the same language combinations and tend to concentrate on purely morphosyntactic properties of the target language.

The present study endeavours to fill some of these lacunae: the language combination which is tested, L1 English/L2 Dutch, is severely underrepresented in the summary of studies given in the summary table in Appendix A, occurring in just one study (Snow and Hoefnagel-Höhle 1982). Furthermore, one of the experimental methods which will be used, the truth value judgement task, has rarely (if ever) been used with L2 children before. Additionally, the target language property which is investigated, direct object scrambling in Dutch, goes beyond morphosyntax – it involves aspects of syntax, semantics and discourse/pragmatics.

It is to scrambling which we now turn.

CHAPTER 2

DIRECT OBJECT SCRAMBLING IN DUTCH

Introduction

A characteristic property of languages such as Dutch, German and Japanese, scrambling – a term coined by Ross in his (1967) dissertation – refers to the reordering of sentential constituents, such as direct and indirect objects, within and/or across clause boundaries. For example, in Dutch (1) and (2), the NP and PP objects, *de tuin* 'the garden' and *voor haar zaden* 'for her seeds', can appear to either the right or the left of the sentential adverbs, *vandaag* 'today' and *nu al* 'now already'. In German (3), the direct object *die größte Brombeere* 'the biggest blackberry' can appear to the right or the left of the indirect object *ihrer Mutter* 'her mother', as well as in pre-subject position, as in (3)-c. In Japanese, scrambling is possible across clause boundaries, as exemplified in (4)-b, where the direct object *kono hanao* 'this flower' appears in the pre-subject position of the matrix clause rather than directly adjacent to the matrix verb, as in (4)-a.

(1)	a.	Willemijn	heeft	vandaag	[de tuin]	omgespit
		Willemijn	has	today	the garden	up-dug
	b.	Willemijn	heeft	[de tuin]	vandaag	omgespit
		Willemijn	has	the garden	today	up-dug

'Willemijn dug the garden today.'

(2)	a.	Elma	heeft	nu	al	[voor haar zaden]	betaald
		Elma	has	now	already	for her seeds	paid
	b.	Elma	heeft	[voor haar zaden]	nu	al	betaald
		Elma	has	for her seeds	now	already	paid

'Elma has paid for her seeds already.'

(3)	a.	Kai	sagte	dass	Verena	ihrer Mutter	[die Brombeere]	geben	würde
		Kai	said	that	Verena	her mother	the blackberry	give	would
	b.	Kai	sagte	dass	Verena	[die Brombeere]	ihrer Mutter	geben	würde
		Kai	said	that	Verena	the blackberry	her mother	give	would
	c.	Kai	sagte	dass	[die Brombeere]	Verena	ihrer Mutter	geben	würde
		Kai	said	that	the blackberry	Verena	her mother	give	would

'Kai said that Verena would give the blackberry to her mother'

(4) a.

Akira-ga	Barbara-ga	[kono-hana-o]	tunda	to	omotteiru
Akira-Nom	Barbara-Nom	this flower-Acc	picked	that	thinks

b.

[Kono-hana-o]	Akira-ga	Barbara-ga	tunda	to	omotteiru
this flower-Acc	Akira-Nom	Barbara-Nom	picked	that	thinks

'Akira thinks that Barbara picked this flower.'

This thesis deals with the acquisition of Dutch, one of the more restricted scrambling languages.[1] This chapter details the properties of scrambling which will be tested in the experimental work presented in subsequent chapters. In particular, §2.1 briefly outlines the characteristic properties of Dutch scrambling in a comparative light, before turning to the specifics of the interpretive differences between scrambled and non-scrambled NP objects, the focus of the present work. Section 2.3 presents a brief review of some recent accounts of Dutch scrambling, and in §2.4, the analysis which is adopted here is given. The non-native (L2) subjects in the present study are all native speakers of English and, hence, it is also necessary to chart the differences between this language and Dutch. This is done descriptively in §2.2 and in theoretical terms in §2.5.

2.1 Properties of scrambling

Compared to other languages, scrambling in Dutch is rather limited. Unlike Japanese, it is clause-bound, as illustrated in (5)-a, and unlike German, it is restricted to the *Mittelfeld*, that is, pre-subject scrambling in the same clause, as in (3)-c, is generally not allowed, as shown in (5)-b.

(5) a.

*Alison	zei	[kruisbessen]	dat	ze	elke dag	zou	kunnen	eten
Alison	said	gooseberries	that	she	every day	would	can	eat

b.

*Alison	zei	dat	[kruisbessen]	ze	elke dag	zou	kunnen	eten
Alison	said	that	gooseberries	she	every day	would	can	eat

Note, however, that scrambling to the pre-subject position becomes possible once the scrambled constituent is heavily focussed, as marked by stress (indicated by SMALL CAPITALS), as in (6)-a, or by focus markers such as *zulke* 'such' and *zelfs* 'even', as in (6)-b (Neeleman 1994).

[1] It is possibly because of its relatively restricted nature that scrambling in Dutch has sometimes been likened to object shift (OS). Dutch scrambling differs from what is generally known as OS in Mainland Scandinavian languages, however, for (at least) the following reasons: (i) scrambling is not dependent on movement of the main verb, whereas OS is; and (ii) both PP and NP objects can scramble, whereas only NP objects shift (see Holmberg 1986; Thráinsson 2001; Vikner 1994 and references cited there). See Thráinnison (2001) for an overview of the differences and similarities between scrambling and object shift.

(6) a.

Alison	zei	dat	[KRUISBESSEN]	ze	elke dag	zou	kunnen	eten
Alison	said	that	gooseberries	she	every day	would	can	eat

'Alison said that she could eat GOOSEBERRIES every day.'

b.

Alison	zei	dat	[zulke	kruisbessen]	zelfs	zij
Alison	said	that	such	gooseberries	even	she

niet	zou	kunnen	eten
not	would	can	eat

'Alison said that even she could eat such gooseberries.'

Similarly, (clause-internally) scrambled indirect NP objects are only considered acceptable when marked in this way:[2]

(7) a.

?*	Henrietta	gaat	[Tom]	niet	de kruiwagen	geven
	Henrietta	goes	Tom	not	the wheelbarrow	give

b.

Henrietta	gaat	[TOM]	niet	de kruiwagen	geven
Henrietta	goes	Tom	not	the wheelbarrow	give

'Henrietta is not going to give TOM the wheelbarrow.'

b.

Henrietta	gaat	[zelfs Tom]	niet	de kruiwagen	geven
Henrietta	goes	even Tom	not	the wheelbarrow	give

'Henrietta is not going to give even Tom the wheelbarrow.'

This 'focus-scrambling' has been argued to differ from 'normal' scrambling (Neeleman 1994) and hence, it will be excluded from consideration in the present study.

As exemplified in (1) and (2) above, both NP and PP objects can scramble. This thesis deals with NP objects only, however. In the present context, the term scrambling should therefore be understood descriptively as the occurrence of an NP object to the left of an adverbial.

Different NP types pattern differently in their scrambling behaviour: definite NPs (8) generally scramble optionally, pronouns (9) scramble obligatorily, and indefinite NPs (10) generally do not scramble, that is, indefinite NPs occur predominantly in non-scrambled position, as in (10)-a; to some native speakers, they sound slightly odd when, as in (10)-b, they are scrambled out-of-the-blue – see below for relevant discussion.

(8) a.

Henk	heeft	gisteren	[de gereedschapskist]	gebouwd
Henk	has	yesterday	the tool box	built

b.

Henk	heeft	[de gereedschapskist]	gisteren	gebouwd
Henk	has	the tool box	yesterday	built

'Henk built the tool box yesterday.'

---

[2] Indirect objects realised as PPs are scrambled freely (Zwart 1997:32).

(9) a.

Ien	heeft	[hem]	al	gezien
Ien	has	him	already	seen

b.

*Ien	heeft	al	[hem]	gezien
Ien	has	already	him	seen

'Ien has seen it already.'

(10) a.

Jos	heeft	gisteren	[een schop]	gekocht
Jos	has	yesterday	a spade	bought

b.

Jos	heeft	[een schop]	gisteren	gekocht
Jos	has	a spade	yesterday	bought

'Jos bought a spade yesterday.'

This thesis looks at the acquisition of the scrambling of definite and indefinite NP objects only. Specifically, it examines whether non-native acquirers (L2ers) are able to acquire the different interpretations which are associated with the different word orders for these two types of object NPs. These differences will be outlined in the following two sections.

2.1.1 Definite NPs

As noted above, definite NPs are generally considered to scramble optionally (e.g. de Hoop 2000). In the examples given in (11) and (12), there is no truth-conditional difference in meaning between the non-scrambled and the scrambled orders.

(11) a.

John	heeft	gisteren	[de boom]	geplant
John	has	yesterday	the tree	planted

b.

John	heeft	[de boom]	gisteren	geplant
John	has	the tree	yesterday	planted

'John planted the tree yesterday.'

(12) a.

Gert Jan	heeft	twee keer	[de bloem]	besproeid
Gert Jan	has	two times	the flower	watered

b.

Gert Jan	heeft	[de bloem]	twee keer	besproeid
Gert Jan	has	the flower	two times	watered

'Gert Jan watered the flower twice.'

Generally, however, there is a tendency for anaphoric definite NPs, that is, definite NPs which refer back to another constituent in the context, to appear in scrambled position and for non-anaphoric definite NPs to appear in non-scrambled position. When negation is involved, the situation is slightly more complex. Although once again, there is no truth-conditional difference between the scrambled and non-scrambled orders, negation and focus interact such that the scrambled order, given in (13)-a, is generally interpreted as expressing sentential negation, and the non-scrambled order, given in (13)-b, is interpreted as contrastive negation, either of the object or of the whole VP.[3] Both sentences are given with neutral stress, indicated by SMALL CAPITALS.

(13)	a.	Bob	heeft	[het onkruid]	niet	WEGgegooid
		Bob	has	the weeds	not	away-thrown
		'Bob didn't throw away the weeds.'				
	b.	Bob	heeft	niet	[het ONKRUID]	weggegooid
		Bob	has	not	the weeds	away-thrown
		'Bob didn't throw away the WEEDS.'				

The results of the production experiment in Chapter 5 will show that when sentential negation is intended, scrambling (of an anaphoric definite NP object) is more or less obligatory for native speakers.

The fact that the non-scrambled order induces a contrastive reading is indicated by the continuation sentences for (13)-b given in (14). When the object is contrasted, the presupposition is that something has been thrown away; (14)-a states that this was the young tomato plants rather than the weeds. When the VP is contrasted, the presupposition is that an event happened in the garden; (14)-b states that this involved hoeing down the rose bushes rather than throwing away the weeds. Note that the scrambled sentence in (13) can also be interpreted as negating the VP. For this reason, I concentrate only on the difference between the scrambled order associated with sentential negation and the non-scrambled order used to express contrastive negation on the object.

(14)	a.	Hij	heeft	de jonge	TOMATENplanten	weggegooid
		he	has	the young	tomato plants	away-thrown
		'He threw away the young tomato plants.'				
	b.	Hij	heeft	de ROZENstruiken	ondergeschoffeld	
		he	has	the rose bushes	under-hoed	
		'He hoed down the rose bushes.'				

With respect to definite NPs, the present study is concerned with the acquisition of scrambling across negation only, that is, it asks whether English-speaking L2ers know that when a sentential rather than contrastive negation reading is intended, the definite NP should be scrambled.

2.1.2 Indefinite NPs

More often than not, indefinite NP objects do not scramble. When they do, however, there is a clear difference in interpretation between the scrambled

[3] *Modulo* heavy stress on the subject indicating otherwise. Note also that a contrastive reading of the object is also possible in (13)-a if the object is heavily stressed. This is marked in comparison with (13)-b, however.

and non-scrambled word orders. This is illustrated below for sentential adverbs (15), VP-adverbs (16) and negation (17).

(15)	a.	Cécile	heeft	gisteren	[een roos]	geplant	
		Cécile	has	yesterday	a rose	planted	
		'Cécile planted a(ny) rose yesterday.'					
	b.	Cécile	heeft	[een roos]	gisteren	geplant	
		Cécile	has	a rose	yesterday	planted	
		'Cécile planted a (certain) rose yesterday.'					
(16)	a.	Simon	heeft	twee keer	[een roos]	besproeid	
		Simon	has	two times	a rose	watered	
		'Simon watered a(ny) rose twice.'					
	b.	Simon	heeft	[een roos]	twee keer	besproeid	
		Simon	has	a rose	two times	watered	
		'Simon watered a (certain) rose twice.'					
(17)	a.	Brigit	heeft	geen	(niet+een)	roos	geplukt
		Brigit	has	no	(not+a)	rose	picked
		'Brigit didn't pick a(ny) rose.'					
	b.	Brigit	heeft	[een roos]	niet	geplukt	
		Brigit	has	a rose	not	picked	
		'Brigit didn't pick a (certain) flower.'					

The indefinite object NP in the non-scrambled orders in (15)-a through (17)-a takes narrow scope with respect to the adverb, that is, it is interpreted within the scope of the adverbial, whereas in the scrambled orders in (15)-b through (17)-b, it takes wide scope, that is, it is outside the scope of the adverbial. Consequently, as indicated by the glosses, the non-scrambled indefinite *een roos* 'a rose' is interpreted as any rose, whereas the scrambled indefinite object NP singles out a particular rose. This is clearest for the scopal operators *twee keer* 'twice' and *niet* 'not' in (16) and (17). A similar effect is observed in (15) for *gisteren* 'yesterday', although here, the semantics are far from straightforward (Ruys 2001). The experiments in Chapters 5 and 6 focuses on the clear-cut cases illustrated in (16) and (17).

The adverb *twee keer* (16) clearly separates two different NP positions. This is not the case, however, when negation is involved, as in (17). The scrambled form in (17)-b alternates with the suppletive *geen* 'no(ne)/not any' form in (17)-a. *Geen* can be analysed as the incorporation of the negator into the indefinite determiner. Given that this incorporation can only take place if the negator occurs directly to the left of the indefinite (Rullman 1995), sentences such as (17)-a are counted as non-scrambled forms (see §5.3.2.2).[4] Further support for this analysis comes from the observation that the *geen* form shares the same non-specific interpretation as the non-scrambled *niet* + *een* form and not the specific interpretation associated with the scrambled form. The use of the separate *niet* and *een* forms in the non-scrambled position is not ungrammatical, but it is highly dispreferred. Furthermore, the non-scrambled *niet* + *een* form can be interpreted contrastively, especially when the object is (heavily) stressed, as in (18)-a. This also the case for the *geen* form when the object is (heavily) stressed, as in (18)-b. That these two sentences involve a contrastive interpretation is illustrated with the continuation sentence in (18)-c.

(18) a. Brigit heeft niet [een RÓÓS] geplukt
Brigit has not a rose picked
'Brigit didn't pick a rose.'

b. Brigit heeft [geen RÓÓS] geplukt,
Brigit has no rose picked
'Brigit didn't pick a rose.'

c. … maar een tulp
… but a tulip
'….rather, (she picked) a tulip.'

Scrambled indefinite object NPs have been variously labelled as 'specific' (in the sense of Enç 1991), 'referential' (following Fodor and Sag 1982), 'presuppositional' (Diesing 1992) or 'strong' (de Hoop 1992). To avoid any confusion, the cover-term 'specific' (in its intuitive sense) will be used throughout this thesis, although, strictly-speaking, the reading which is tested in the relevant experimental conditions is partitive, in De Hoop's (1992) terms, or 'source-set', on Krämer's (2000) analysis, for example.

Scrambled indefinite objects thus receive a specific interpretation and non-scrambled indefinites a non-specific interpretation. Note that the entailment relations between the scrambled and non-scrambled sentences vary depending on which scopal operator is involved. The scrambled sentence in (16)-b entails the non-scrambled sentence in (16)-a, because if there is a rose which was watered twice, there are also two rose-watering events. This does not hold the other way round, because if there were two rose-watering events, it is not necessarily the case that they involve one and the same rose. The situations which verify the non-scrambled sentence therefore include those which verify the scrambled sentence. In Chapter 3, I will show that this fact has important implications for how the interpretive properties of scrambled indefinites are acquired. When negation is involved, as in (17), the opposite

[4] Given that there is usually an adjacency requirement on incorporation (see e.g. Van Geenhoven 1998), it is not clear how (17)-a could be analysed if the object were underlyingly scrambled.

holds. There, the non-scrambled sentence in (17)-a entails the scrambled sentence in (17)-b, because if no rose was picked, there must (on the assumption that the set of roses in the model is not empty) be a rose such that it was not picked. This does not hold the other way round, because if there was one rose which was not picked, then it is not necessarily the case that no roses were picked at all. This means that the situations which make the scrambled sentence true include those which make the non-scrambled sentence true. This will have important consequences for the set-up of the comprehension experiments in Chapter 6.

The indefinite article, *een*, in Dutch is generally pronounced in its unstressed form, /ən/ (written as *een* or '*n*), but it can also be pronounced as /e:n/ (written as *één*). The status of *één* is rather unclear. Traditionally, *één* is analysed as the cardinal *one* (Haeseryn 1997), but it has also been argued that because *één* does not always behave like other cardinals, it should be analysed as the stressed form of the indefinite determiner (Barbiers 2005). For example, *één* can form the basis for a nominal derivation with the suffix *–heid* '-ity', whereas cardinals such as *twee, drie, tien etc.* 'two, three, ten, etc' cannot (compare *éénheid* with **tweeheid*, **drieheid*, **tienheid*, etc.), and whereas the latter can combine with the derivational suffix *–tal* 'number', *één* cannot (compare *tweetal, drietal, tiental* etc. with **ééntal*). Furthermore, *één* can be used predicatively, whereas *twee, drie, tien,* etc. cannot:

(19)	Man	en	vrouw	zijn	één / *twee
	man	and	woman	are	one / *two
	*'Man and woman are one/*two.'*				(Barbiers 2005:165, ex. 111)

In this thesis, I will remain agnostic as to whether *één* is a cardinal or a stressed indefinite and I refer to it as the cardinal/stressed indefinite throughout (see §5.4.3.3, §6.3.3.3, §6.4.3.1 and §7.4.2 for further discussion).

When asked about their intuitions, some native speakers prefer to hear/produce the cardinal/stressed form, *één*, in scrambled position (possibly to emphasise the singularity of the scrambled object). However, regardless of this preference, for all speakers questioned, it is the (scrambled) position of the indefinite alone which, when presented in the relevant context, is enough to indicate that a specific reading is intended, irrespective of whether the indefinite article is realised as *een* or *één*. We shall see that the experimental results presented in Chapters 5 and 6 confirm these intuitions.

2.2 Properties of the L1: English

This thesis investigates the L2 acquisition of direct object scrambling in Dutch by native English-speaking children and adults. In addition to the properties of scrambling, it is therefore necessary to consider the relevant properties of the L1, English. English does not allow scrambling, that is, there is no free word order between adverbials and direct objects with concomitant interpretive effects as in Dutch. This is illustrated in (20) through (22).[5]

(20) a. Cécile planted [a rose] yesterday
b. *Cécile planted yesterday [a rose]
c. *Cécile yesterday planted [a rose]
d. Yesterday, Cécile planted [a rose]

(21) a. Simon watered [a rose] twice / two times
b. *Simon watered twice / two times [a rose]
c. Simon ?twice / *two times watered [a rose]
d. Twice, Simon watered [a rose]

(22) a. Brigit has not picked [a rose]
b. *Brigit has not [a rose] picked
c. *Brigit has [a rose] not picked
d. *Not Brigit has picked [a rose]

Adverbs in English cannot appear between the verb and the direct object. Rather, depending on the type of adverbial in question, they appear in one or more positions: sentence-initially, sentence-finally and/or directly following the subject.[6]

Note that in English, sentences containing NPs with determiners such as *a*, *one*, *three*, etc., are ambiguous with respect to specificity. In each of the examples in (23), the indefinite object NP could – in a suitable context – refer to one rose or to any rose.

---

[5] Johnson (1991) argues that English has ("Scandinavian") object shift, that is, that both the verb and the object always move to the specifier position of some higher functional projection (see also Koizumi 1993; Takano 1998). This, he claims, is visible in the particle construction, illustrated in (i).

(i) a. Mikey looked [the reference] up
b. Mikey looked up [the reference]

(Johnson 1991:593, ex. 541)

Following Collins and Thráinsson (1996, amongst many others), I will assume that it is the particle which moves, rather than the object (see also the discussion of Gilkerson (2004) in §1.4.3.2 on the acquisition of particle verbs in English). Note, in any case, that even if one were to assume an analysis of English where the object always moves to, for example, [Spec, AgrOP], this movement is not associated with the interpretive differences which are found between scrambled and non-scrambled orders in Dutch – see below for further discussion. (Under certain circumstances, however, such interpretive differences are observed in English, namely in the double-object construction with an indefinite indirect object and a universally quantified direct object, as in *Sydney gave a bird every crumb*. This is discussed more fully in Chapter 3 (§3.4, fn. 27), where I show that it is not relevant in the context of the present study.)

[6] Judgements regarding *twice* vary across speakers and across varieties. As indicated by the diacritics in (21), not all speakers of British English allow *twice* to appear in directly post-subject position and with the American English equivalent, *two times*, this is impossible.

(23) a. Cécile planted [a/one rose] yesterday
b. Simon watered [a/one rose] twice
c. Brigit didn't pick [a/one rose]

That the sentences in (23) are ambiguous is illustrated for (23)-b, *Simon watered a flower twice*, with the continuation sentences in (24).

(24) a. It grew very quickly
b. They both bloomed afterwards

The sentence in (24)-a indicates that the object can have a specific reading (that is, that there was a certain flower which Simon watered twice) and the sentence in (24)-b shows that a non-specific reading is possible (that is, that Simon took part it two flower-watering events).[7]

There are a limited number of ways to unambiguously realise a specific reading in English. The first is to use an indefinite article in conjunction with an adjective such as *particular* or *certain*, as illustrated in (25) (see Enç 1991 for relevant discussion).

(25) a. Cécile planted [a certain rose] yesterday
b. Simon watered [a certain rose] twice
c. Brigit didn't pick [a certain rose]

Another is to use the plural indefinite determiner *some*, which does not allow a non-specific interpretation when embedded under negation, as in (26).

(26) Berit didn't pick [some roses]

The specific reading is also the strongly preferred – for some speakers, the only – interpretation for the overt partitive in (27) (see Musolino and Lidz to appear for relevant discussion).

(27) a. Cécile planted [one of the roses] yesterday
b. Simon watered [one of the roses] twice
c. Brigit didn't pick [one of the roses]

The sentences in (25) are more likely to be used when the speaker has a particular object in mind, whereas the overt partitive in (27) makes a clear link to the preceding discourse. This is not necessarily the case for the indefinite *some Ns* or the indefinite *a N* in combination with *certain* or *particular*.

To summarise: In contrast to Dutch, English does not have scrambling and there are limited means to unambiguously realise the specific reading of an indefinite object NP, namely by using the overt partitive or an adjective such as

[7] Further evidence that sentences such as *Brigit didn't pick a rose* are ambiguous comes from experimental work by Miller and Schmitt (2004) and Su (2001). This is discussed in §3.3.1.3.

certain or *particular* (both of which, in addition to scrambling, are also possible in Dutch), or in negative sentences by using the plural indefinite determiner *some*.

Having seen the descriptive facts about scrambling, we now turn to how this property of Dutch has been analysed in the theoretical literature.

2.3 Approaches to scrambling

The topic of scrambling has generated much debate in the theoretical literature (see Bailyn 2002; Corver and van Riemsdijk 1994a; Grewendorf and Sternefeld 1990; Thráinsson 2001 for overviews). There is, however, little consensus as to what the precise analysis of scrambling should be. Earlier work concentrated on categorising scrambling as A- or A-bar movement (or both) (see some of the papers in Corver and van Riemsdijk 1994b, for example), whereas in recent years – and particularly from a minimalist perspective (Chomsky 1995) – it has been the optionality of scrambling which has been the focus of attention. One of the issues in this debate has been the question of whether scrambling is 'triggered' by a particular feature or property associated with the object. Numerous solutions to this question have been proposed, including analyses where objects scramble because they have certain semantic and/or syntactic properties, approaches where the semantic and discourse properties of scrambled objects *follow* from scrambling rather than trigger it, and yet other analyses where scrambling is not triggered by anything at all.

A detailed presentation and evaluation of all of these different analyses and the issues these give rise to is beyond the scope of this thesis. Therefore, the present section only focuses on approaches: (i) which deal specifically with Dutch (and preferably also compare this language with English); (ii) which claim to account for the interpretive differences between scrambled and non-scrambled indefinite NPs; and/or (iii) which are adopted by the researchers whose approaches to the (L1) acquisition of scrambling are important in subsequent chapters, namely Krämer (2000) and Schaeffer (2000). Each approach will be evaluated with respect to how it explains the syntactic and semantic properties of scrambling, that is, how the different word orders (scrambled and non-scrambled) are generated and how the different interpretations are derived. It is crucial to recognise that none of the instances of scrambling which are investigated in this thesis is truly optional. Hence, the much-discussed issue of the optionality of scrambling will, to a great extent, be side-stepped here.

Note that it is not a goal of the present work to evaluate the different approaches to scrambling using acquisition data (cf. Krämer 2000). With respect to the final analysis of scrambling which I adopt, I endeavour to remain as analysis-neutral as far as is possible. Where particular approaches do not differ in their empirical coverage or their conceptual elegance, one will not be chosen over the other. Such a method is adopted so that any conclusions drawn in the present study do not crucially depend on the technicalities of a particular syntactic (or semantic) theory, which, if revised, may demand very different conclusions. As Schwartz and Sprouse (2000) point out, any conclusions based on L2 research which is deeply embedded in the technicalities of one particular theory are only as provisional as the theory itself. This point is especially pertinent here, because, as indicated above, there is little consensus as to how scrambling can best be analysed. Nevertheless, a precise theoretical framework is an indispensable tool for any (generative) study of interlanguage grammars (Schwartz and Sprouse 2000); the analysis which will be adopted here is outlined in §2.4.

Although analyses of scrambling must – due to the nature of the beast – incorporate information from several different components of language (including syntax, semantics, pragmatics, discourse), they differ considerably with respect to the emphasis they place on these components. Some see scrambling as an effect of the properties of particular object NPs (e.g. de Hoop 1992), sometimes formally incorporating these properties into the syntax (e.g. Schaeffer 2000). For others, the variation in word order exhibited in Dutch falls out from some other (non-interpretive-related) property of grammar and the interpretive differences are accounted for by other means (e.g. Neeleman and Reinhart 1998; Zwart 1997). The syntactic, discourse(-related) and semantic/pragmatic approaches to scrambling are presented in §2.3.1, §2.3.2 and §2.3.3, respectively.

2.3.1 Syntactic approaches to scrambling

Traditionally, scrambling in Dutch has been analysed as an instance of A-movement for Case-checking purposes, largely as a result of the properties which it shares with other types of A-movement (Vanden Wyngaerd 1989). For example, it creates new binding possibilities, blocks other binding possibilities and does not yield weak crossover effects. (See Thráinsson 2001 for an overview of the standard tests which are employed.).[8] This section examines two approaches to scrambling which develop Vanden Wyngaerd's original insight in light of Kayne's (1994) anti-symmetry hypothesis. Zwart (1997) and Koster (1999; 2000) assume that Dutch is underlyingly VO and that scrambling is Case-driven movement which serves, in part, to derive the OV word order in Dutch observable in embedded clauses and participal constructions. The different word order options are derived from the free adjunction of adverbs and the interpretive differences associated with these word orders are analysed either as a reflex of the particular properties of adverbs and the different object NPs (Zwart 1997) or as a consequence of linear scope (Koster 1999; 2000). Both of these analyses clearly outline the

---

[8] Bennis and Hoekstra (1984) argue that Dutch scrambling licenses parasitic gaps and must therefore be A-bar movement. Neeleman (1994) shows, however, that this argument does not hold (see Haider and Rosengren 2003 for further arguments against scrambling licensing parasitic gaps).

differences between Dutch, on the one hand, and English, on the other, and for this reason, are included for review here.

2.3.1.1 Zwart (1997)

For Zwart, Dutch and English differ in that direct objects in Dutch always move to [Spec, AgrOP] to check Case overtly, whereas in English, this Case-checking takes place covertly. In other (minimalist) words, the morphological features in AgrO which force the object to move are strong in Dutch and weak in English. On the basis of alternations such as (28), adverbs are assumed to be generable in various positions:

(28)	a.	…dat	gisteren	Jan	[Marie]	gekust	heeft	
		…that	yesterday	John	Mary	kissed	has	
		'…that John kissed Mary yesterday.'						
	b.	…dat	Jan	gisteren	[Marie]	waarschijnlijk	gekust	heeft
		…that	John	yesterday	Mary	probably	kissed	has
		'…that Jan probably kissed Mary yesterday.'						

(Zwart 1997:64, ex. 107)

Unlike in the other approaches discussed below, the interpretive properties of non-scrambled indefinite object NPs do not result from the object's VP-internal position. On Zwart's approach, this is quite simply because no object NP remains in this position in Dutch, because they all have to move to check their Case feature. The specific interpretation of scrambled indefinite objects is basically viewed as a consequence of the relationship between stress (accent) and focus. When indefinite object NPs are unstressed (deaccented), they are interpreted as old information, and this leads to a specific interpretation. For Zwart, the fact that deaccented (specific) indefinite objects appear to the left of adverbs (in the position which has hitherto been referred to as the scrambled position) and that accented (non-specific) indefinite objects appear to their right (in non-scrambled position) falls out of 'proper generalisations regarding the placement of adverbs' (1997:96). Specifically, he claims that the distributional properties of indefinite object NPs are a consequence of the particular properties of the different kinds of adverbs: sentence adverbs (such as *gisteren* 'yesterday', *altijd* 'always'), VP-adverbs (*snel* 'quickly', *hard* 'loudly') and modal particles (*maar* '(non-temporal) just/only', *even* (lit.) 'for a while'). Here, I discuss the case of sentence adverbs.

Zwart claims (but does not illustrate) that stressed indefinite NPs prefer to appear to the right of sentential adverbs for two reasons. First, when they receive a secondary pitch accent, indicated in (29) with a grave accent (`), these adverbs may behave like VP-adverbs. VP-adverbs constitute their own prosodic domain. If an object were to appear to the left of a VP-adverb – or, by extension, to the left of the sentence adverb *gisteren* 'yesterday' – the sentential adverb would be nested in the prosodic domain created by the object and the verb and such nesting is not allowed (Truckenbrodt 1995, cited in Zwart 1997).

(29)	a.	Jan	heeft	gìsteren	[een BOEK]	gelezen
		John	has	yesterday	a book	read
		'John read a BOOK yesterday.'				
	b.	*Jan	heeft	[een BOEK]	gìsteren	gelezen
		John	has	a book	yesterday	read

Second, when sentence adverbs are unstressed, as in (30), they are unable to be in focus and consequently they must appear to the left of the stressed indefinite object NP.

(30)	a.	Jan	heeft	gisteren	[een BOEK]	gelezen
		John	has	yesterday	a book	read
		'John read a BOOK yesterday.'				
	b.	*Jan	heeft	[een BOEK]	gisteren	gelezen
		John	has	a book	yesterday	read

Note that without any context, these judgements are quite tricky. Note, furthermore, that with a secondary pitch accent on the verb, (30)-b is rendered acceptable. On Zwart's analysis, then, the reason why stressed indefinite objects generally occur to the right of adverbs is a result of the properties of these adverbs, and the reason why indefinite objects to their left receive a specific interpretation is because they are deaccented and therefore refer to old information.

There are, however, (at least) three problems with the link made between stress and meaning in Zwart's analysis. Firstly, it cannot account for why a specific reading is still available when an indefinite object to the left of the adverb, as in (31), is *not* unstressed. The sentential adverb in (31) does not receive a secondary pitch accent, but unlike (29)-a, the stressed indefinite object appears to its left rather than to its right. Furthermore, it is not in focus because it is unstressed. Nevertheless, the indefinite object is scrambled and it has a specific reading, that is, there is a particular rabbit which Carolyn saw.

(31)	Carolyn	heeft	[een KONIJN]	gisteren	in de tuin	gezien
	Carolyn	has	a rabbit	yesterday	in the garden	seen
	'Carolyn saw a RABBIT in the garden yesterday.'					

Secondly, Zwart's analysis cannot account for why the specific reading is unavailable when the verb is stressed rather than the non-scrambled object, as in (32). If the deaccenting of an indefinite object is supposed to mean that it is old information and this leads to a specific interpretation, then the object in

(32) should have a specific interpretation, but it does not; it has a non-specific interpretation (as in (30)-a).[9]

(32)	Julia	heeft	gisteren	een bloem	moeten	beSPROEien
	Julia	has	yesterday	a flower	must	water
	'Julia had to water a (non-specific) plant yesterday.'					

A third problem with Zwart's account is that it cannot explain why indefinite objects which appear to the left of the adverbs (i.e. scrambled indefinite objects) can have different readings (see §2.3.3.1 for further discussion of this point). Taken together, these three problems suggest that although focus and intonation clearly play a role in direct object scrambling in Dutch, the properties of adverbs and the interaction between intonation and focus cannot be the entire story when it comes to accounting for the interpretive effects which are observed (see also discussion in §2.3.2.2).

2.3.1.2 Koster (1999; 2000)

Koster (1999; 2000) offers, at first sight, an elegant approach to scrambling in Dutch. It is appealing from an L2 acquisition perspective because by specifically addressing the differences between Dutch and English, it reduces the acquisitional task for the English-speaking L2er of Dutch to the resetting of just one parameter. The proposal works as follows. All VP-internal material is licensed by overt movement to some functional projection (AgrSP, Dat(ive)P, Acc(usative)P) to the left of the VP in the following structure:

(33) …[AgrSP [Adv1* [T [Dat [Acc [Adv2* [[… [VP]]]]]]]]]]

Direct objects always move to [Spec, AccP]. The position of the object in relation to the adverb depends on which type of adverb is present. There are two classes of adverbs: Adv1, the type of adverb which can (but need not) precede the VP in English, such as *probably*, and Adv2, the type of adverb

---

[9] Zwart (1995) discusses a similar example with bare plural indefinites. In (i), the stress has been moved from the object in (i)-a to the verb in (i)-b. According to Zwart, this results in a shift in interpretation, namely from existential to generic.

(i)	a.	…dat	de politie	altijd	illeGAlen	arresteert
		…that	the police	always	illegal aliens	arrests
		'…that the police is always arresting illegal aliens.'				
	b.	…dat	de politie	altijd	illegalen	arresTEERT
		…that	the police	always	illegal aliens	arrests
		'…that what the police always does to illegal aliens is arrest them.'				

This, he claims, is evidence that it is the stress pattern of the sentence rather than its word order which determines interpretation. This does not hold for (32), however.

which cannot precede the VP in English, such as *yesterday*. These appear optionally in the positions marked with a kleene-star in (33).

On analogy with pied-piping of *wh*-phrase, these two languages are claimed to differ in the size of the phrase in which checking takes place. English allows either the whole *wh*-phrase or a part thereof to move to check the relevant feature (34), whereas Dutch only allows the whole *wh*-phrase to move (35).

(34) a. *Who* did you talk [$_{PP}$ with *t*]?
b. [$_{PP}$ *With whom*] did you talk *t* ?

(35) a. **Wie* heb je [$_{PP}$ met *t*] gepraat?
who have you with spoken

b. [$_{PP}$ *Met wie*] heb je *t* gepraat?
with who(m) have you spoken
'Who did you speak to?'

(Koster 2000:8, ex. 36 and 37)

Along these lines, Koster proposes the 'Pied-Piping parameter', as follows:[10]

(36) Pied-Piping parameter
Dutch checks its VP-internal constituents *individually* (by moving them separately), English *collectively* (by moving the whole VP)

(Koster 2000:8, ex. 38)

Recall that all VP-internal material is licensed by overt movement to a functional projection left of the VP. Koster's proposal is that in English, the VP moves through all Spec positions and the relevant features are checked off in these positions after percolating up via the VP, whereas in Dutch, the object moves separately to the position where its (Case) features are checked off. Accordingly, in essentially the same way as Zwart, scrambling on Koster's account can be seen as a reflex of the object moving to [Spec, AccP] to check Case and the different adjunction sites available for adverbs. The syntactic structure (with only the necessary projections) for non-scrambled and scrambled objects is as follows:

[10] Note that at this point, the analogy with pied-piping breaks down: English allows constituents in a *wh*-phrase to move *either* individually *or* collectively. In the Pied-Piping parameter given in (36), languages choose *between* these two options, that is, they do not (and presumably cannot) allow both. Furthermore, both English and Dutch appear to have opposite settings for Pied-Piping with PPs and Pied-Piping with VP-internal constituents.

(37) a. [$_{AgrSP}$ Heather [$_{T}$ heeft [$_{Adv1}$ twee keer [$_{Acc}$ een goudsbloem$_i$ [$_{VP}$ besproeid t_i]]]]]
Heather has two times a marigold watered
'Heather watered a(ny) marigold twice.'

b. [$_{AgrSP}$ Heather [$_{T}$ heeft [$_{Acc}$ een goudsbloem$_i$ [$_{Adv2}$ twee keer [$_{VP}$ besproeid t_i]]]]]
Heather has a marigold two times watered
'Heather watered a (particular) marigold twice.'

On Koster's account, English does not have scrambling in the sense that the VP moves to [Spec, TP] collectively and as a consequence, the object cannot move individually. In this language, the underlying structure would be as in (38). The adverb *twice*, like its Dutch counterpart *twee keer*, is a type 1 adverb, that is, it can (but need not) precede the VP (at least for some speakers of British English). This is shown in (38)-a and (38)-b; (38)-c is ruled out because *yesterday* is not of the Adv1 class.

(38) a. [$_{AgrSP}$ Heather [$_{T}$ [$_{VP}$ watered a marigold] $_i$ [$_{Acc}$ t_i [$_{Adv2}$ twice [$_{VP}$ t_i]]]]]
b. [$_{AgrSP}$ Heather [$_{Adv1}$ twice [$_{T}$ [$_{VP}$ watered a marigold] $_i$ [$_{Acc}$ t_i [$_{VP}$ t_i]]]]]
c. *[$_{AgrSP}$ Heather [$_{Adv1}$ yesterday [$_{T}$ [$_{VP}$ watered a marigold] $_i$ [$_{Acc}$ t_i [$_{VP}$ t_i]]]]]

Koster (1999; 2000) claims that the interpretive properties of a scrambled object are a matter of linear scope.[11] This is relatively straightforward for the Dutch examples given in (37). In (37)-a, the object *een goudsbloem* 'a marigold' is in the scope of the adverb *twee keer* 'twice'. On the non-specific reading, the sentence can be paraphrased as follows: 'there are two events when Heather watered a marigold'. Whether this was one and the same (thirsty) marigold is irrelevant. In (37)-b, the object is outside of the scope of the adverb and thus, on this specific reading, the sentence is paraphrased as: 'it is the case that for a (particular) marigold x, the number of occasions that Heather watered *x* is two'. In order for this proposal to work for English, however, reconstruction is necessary. Recall that in English, both specific and non-specific readings are available for sentences such as those in (38)-a and (38)-b. In (38)-a, *twice* has scope over a constituent containing the object, namely the trace of the VP, which, on the assumption that reconstruction is allowed, could account for the non-specific ('any marigold') reading; and in its surface position, the object is outside the scope of *twice*, which would derive the specific ('a particular marigold') reading. For (38)-b, the non-specific reading could be read off the surface scope; and to derive the specific reading, Koster would have to say that the object must be insensitive to scope in some way (e.g. Fodor and Sag 1982 – see § 2.5.2 for more details).

There are (at least) three problems with Koster's (1999; 2000) analysis. Firstly, as noted above, without invoking an additional mechanism, that is,

---

[11] According to Koster (1999; 2000), scope is linear if ' "*A* has scope over *B*" means both that *A* c-commands *B* and that *A* precedes *B*' (1999: 14).

reconstruction (which would require some motivation given that, by Koster's definition, scope is always linear – see fn. 11), the differences between Dutch and English cannot be accounted for.[12] Furthermore, and perhaps more importantly, if reconstruction is allowed in English, an additional explanation is required for why it is *not* allowed in Dutch, that is, why can the scrambled object in (37)-b not reconstruct with the consequence that a non-specific reading becomes available? Thirdly, the division of adverbs into two different classes, based on English, seems rather descriptive. Some independent motivation for this is required. It would appear that, given that most adverbs in Dutch occur freely in scrambled and non-scrambled orders (but see the restrictions outlined above in the discussion of Zwart (1997)), most adverbs in Dutch must belong to both classes, which arguably undermines the division in the first place.

2.3.1.4 Summary

In this section we have considered two (predominantly) syntactic analyses of scrambling. Although they both provided a clear account of the relevant differences between English and Dutch, they were both found to be inadequate in some way: Zwart claimed that the interpretive effects of scrambling fall out of the properties of adverbs and intonation, but this was found to be insufficient to account for the relevant data. For Koster, it is linear scope which shoulders the explanatory burden. Without any additional modification, however, his proposal could not adequately account for the English facts. Both of the approaches discussed in this section relate scrambling, at least in part, to Case-checking. This thesis is not concerned with the scrambling of PPs, but they are relevant in this regard. PPs can scramble and they are not assigned Case (Neeleman 1994). A different analysis would

---

[12] Note that a comparable problem exists for some of the data which Koster (1999; 2000) claims his analysis can explain. According to Koster, English exhibits a 'scope paradox': on a traditional analysis, where adverbials are right-adjoined to the VP, higher c-commanding adverbials sometimes take wide scope, as in (i)-a, and sometimes they take narrow scope, as in (i)-b.

(i)	a.	He [[$_{VP}$ worked hard] yesterday]	(yesterday > hard)
	b.	He [$_{VP}$ saw Bill [on his birthday] twice]	(on his birthday > twice)

In other words, scope in English is sometimes linear and sometimes not. On Koster's account, which does not allow right adjunction, the VP in (i)-b has moved in its entirety from its base position to a higher position, allowing the PP *on his birthday* to take scope over the adverbial *twice*, in much the same way as the object in (38). Note, however, that in (i)-b, the alternative scope order (twice > on his birthday) is also possible. To account for this other reading, some additional mechanism would be needed. Koster (2000:6) acknowledges that this other reading exists, but he chooses to ignore it, claiming that it should fall out from another property of English (what he dubs the 'mirror effect'). To go into detail on this matter here would take us too far afield, but note that Koster (1999:38) actually states that 'VP-internal material always has narrower scope than adverbials to the right of the VP'. Clearly, this is just plain wrong.

thus be required in order to derive the correct word orders with PP objects and this would further complicate the approach as a whole.

In the next section, we turn to two discourse-related approaches to scrambling.

2.3.2 Discourse-related approaches to scrambling

In this section, we review two very different approaches to scrambling which can be broadly categorised as 'discourse-related'. Schaeffer (2000) incorporates discourse-based features relating to scrambling into the syntactic derivation, whereas Reinhart (1995) adopts an interface-economic approach, linking the possibility of scrambling to the object NP's status in the discourse. We discuss each of these approaches in turn.

2.3.2.1 Schaeffer (2000)

Building on a feature-driven analysis put forward by Sportiche (1992), Schaeffer proposes that the semantic notion of referentiality, as defined in (39), can be syntactically marked on NPs by means of a [referential] feature.[13]

(39) *Referentiality*: A nominal expression is understood to be referential if it has a "fixed referent" in the (model of the) world, meaning that it can be identified by the speaker and/or by one of the people whose propositional attitudes are being reported.

(Schaeffer 2000:24, ex. 10, based on Fodor and Sag 1982)

Schaeffer distinguishes between discourse-related referential NPs, on the one hand, and non-discourse-related referential NPs, on the other. Discourse-related NPs are NPs which have been introduced in the immediately preceding discourse, whereas non-discourse-related NPs are NPs which are part of any long-term shared knowledge between speaker and hearer, e.g. *the sun*, *the cat* (in a household with a cat), *the Koran*. She observes that these different types of referential (i.e. specific) objects behave differently in relation to sentential adverbs, as illustrated below:

(40) A: Waarom denk je dat …
why think you that
'Why do you think Colette got up so early this morning?'

… Colette zo vroeg is opgestaan vanmorgen?
… Colette so early is up-stood this morning

---

[13] This feature is located in D°; Schaeffer assumes the following internal structure for referential objects:

(i) $[_{DP}\ [_{NumP}\ [_{NP}\ ]]]$

B:	a.	Ik	denk	dat	ze	misschien	[de zon]	wilde	zien	opkomen
		I	think	that	she	perhaps	the sun	wanted	see	up-come

'I think that she perhaps wanted to see the sun come up.'

b.	*Ik	denk	dat	ze	misschien	[het boek]	wilde	uitlezen
	I	think	that	she	perhaps	the book	wanted	out-read

'I think that she perhaps wanted to finish reading the book.'

c.	??Ik	denk	dat	ze	[het boek]	misschien	wilde	uitlezen
	I	think	that	she	the book	perhaps	wanted	out-read

(Schaeffer 2000:36, ex. 30)

In (40)-a, the non-discourse-related NP *de zon* 'the sun' appears to the right of the sentential adverb *misschien* 'perhaps'. This position is not available for the discourse-related NP *het boek* 'the book' in (40)-b. This sentence sounds better when the NP appears to the left of the adverb, but even here, it sounds a little odd, as indicated by the question marks in (40)-c (Schaeffer's judgement). Schaeffer points out that this is simply because there is no antecedent for this NP in the context provided by the question, that is, this NP is not actually discourse-related.

Schaeffer assumes the following syntactic structure:[14]

(41) $[_{CP}$ $[_{AgrSP}$ $[_{Disc(ourse)P}$ (sentential adverb) $[_{TP}$ $[_{Ref(erentiality)P}$ $[_{NegP}$ $[_{AgrOP}$ (VP adverb) $[_{VP}$]]]]]]]]

Scrambled indefinites can presumably only land in [Spec, DiscP], because, by virtue of their indefiniteness, they cannot be referential in the same non-discourse-related sense as non-discourse-related referential definite NPs, such as *de zon* 'the sun'. The fact that the indefinite determiner is often realised as the cardinal/stressed indefinite *één* 'one' (cf. §2.1.2 above) and that this (according to Schaeffer) improves the grammaticality of sentences containing scrambled indefinite objects is indicative of a partitive reading for the indefinite. This in turn shows that scrambled indefinites are discourse-related, thereby, she claims, motivating the analysis that they should appear in [Spec, DiscP]. This is only visible when the object scrambles across a sentential adverb, however, because if an object scrambles across negation or a VP-adverb, it can land in either [Spec, RefP] or in [Spec, DiscP], as outlined above. If an object scrambles across a high adverb, however, it has only one available landing site, namely [Spec, DiscP]. This predicts that non-discourse-related NPs, which according to Schaeffer, can only land in [Spec, RefP] (when they are truly used in this non-discourse-related sense), can scramble over negation and VP-adverbs but they cannot scramble over sentential adverbs. This prediction is not borne out, however. As the example in (42) shows, there are cases where non-discourse-related NPs *can* scramble over high adverbs.

[14] ReferentialityP is the equivalent of Sportiche's (1992) CliticVoiceP.

Although Schaeffer (2000:39) shows that NPs which are usually non-discourse-related can sometimes be used as discourse-related NPs in certain contexts, this is clearly not the case here. The object *de zon* 'the sun' is used out-of-the-blue, in its long-term shared-knowledge sense.

(42) Superman: Volgens mij is Lex Luthor iets boosaardigs van plan
according to me is Lex Luthor sthg terrible from plan
'I think that Lex Luthor is planning something terrible.'

Lois: Ik denk dat hij [de zon] misschien wil opblazen
I think that he the sun perhaps wants up-blow
'I think that maybe he wants to blow up the sun.'

The acceptability of (42) undermines the division Schaeffer makes into discourse-related and non-discourse-related NPs.

One further problem with feature-driven approaches such as Schaeffer's, as Hopp (2003) points out, is that they are inherently contradictory because they require features with interpretive content, such as (non-)discourse-related, to drive movement. On a minimalist approach, only *un*interpretable features are strong and consequently give rise to overt movement.

2.3.2.2 Reinhart (1995), Neeleman and Reinhart (1998)

Whereas Schaeffer attributes the definiteness effects of scrambling to properties of the NP object, Reinhart (1995) and Neeleman and Reinhart (1998) seek to account for these at the level of the PF interface. Specifically, Reinhart (1995) proposes that definite and specific indefinite NP objects appear in a scrambled position because here, they will not receive main stress and, consequently, they will not be in focus. As explained in more detail below, this can be viewed as the equivalent of 'stress-shifting' in English.

Reinhart (1995) follows Cinque's (1993) theory of sentence stress and focus. Cinque claims that sentence stress is assigned to the most deeply embedded constituent in a sentence. Thus, in the English and Dutch examples given in (43)-a and (43)-b, respectively, it is the object which is the most deeply embedded constituent and which is consequently assigned main sentence stress (as indicated by SMALL CAPITALS).[15]

[15] Such an analysis of focus crucially relies on the assumption that there exists a difference between neutral and marked sentence stress. Reinhart provides extensive justification for this assumption, but this will not be dealt with here, except to note that when a sentence has marked stress, it does not necessarily imply that it sounds 'odd' in some way. In fact, it should sound perfectly normal. The reason it is marked is because it has been derived using some marked operation, such as stress-shifting (see Reinhart (1995:40ff.) and Neeleman and Reinhart (1998:338ff.)).

(43) a. [Laurence [raked [the SOIL]]]

b. [Laurence [heeft [de AARDE] geharkt]]
Laurence has the soil raked
'Laurence raked the soil.'

Sentence stress, assigned by independent phonological rules, interfaces with the discourse/context in which the sentence is uttered via the notion of focus. Focus is defined as the most informative part of an utterance. Reinhart claims that an utterance is associated with a set of possible foci. This focus set, that is, those constituents which could be in focus, is defined as follows:

(44) The focus set of IP consists of the constituents containing the main stress of IP

For the Dutch sentence in (43)-b, this would mean that in a sentence with neutral stress, the object, the VP and the IP could be in focus, because each of these constituents contains the main stress of the IP, as exemplified in (45)-a. When the verb carries the main stress, as in (45)-b, the object can no longer be in focus, but the verb can. The same pattern holds for English.

(45) a. [$_{IP}$ Subject [$_{VP}$ OBJECT Verb]]
Focus set: {Object, VP, IP}

b. [$_{IP}$ Subject [$_{VP}$ Object VERB]]
Focus set: {V, VP, IP}

Which of these constituents is selected as the actual focus of an utterance is determined at the interface, that is, it depends on the context in question. This is illustrated in (46) for English, using various questions to elicit the intended focus set. The focussed constituent is enclosed in square brackets.

(46) a. *Focus = IP*
Speaker A: What's the noise?
Speaker B: [My neighbour is building a SHED]

b. *Focus = VP*
Speaker A: What's your neighbour doing?
Speaker B: My neighbour is [building a SHED]

c. *Focus = Object*
Speaker A: What's your neighbour building?
Speaker B: My neighbour is building [a SHED]

d. *Focus = Verb*
Speaker A: Has your neighbour bought a shed already?
Speaker B: No, my neighbour is [BUILDING] a shed

Reinhart applies this theory of sentence stress and focus to direct object scrambling in Dutch as follows. In a non-scrambled sentence, such as (47)-a, the main stress falls on the most deeply embedded constituent, the object *de bonen* 'the beans', and as indicated in (45)-a, the focus set includes the object, the VP and the IP.

(47) a. Ellie gaat vandaag [de BONEN] oogsten
Ellie goes today the beans harvest
'Ellie is going to harvest the BEANS today.'
Focus set: {Object, VP, IP}

b. Ellie gaat [de bonen] vandaag OOGSTEN
Ellie goes the beans today harvest
Ellie is going to HARVEST the beans today.'
Focus set: {Verb, VP, IP}

In the scrambled equivalent, given in (47)-b, the main stress no longer falls on the object, as this is no longer in the most deeply embedded position. That is the verb. Consequently, the focus set includes the verb, the VP and the IP, as in (45)-b.

Cinque (1993) argues that stress shift involves two distinct processes: destressing a stressed element and strengthening an element that does not bear the main stress. Reinhart (1996) expands on Cinque's original distinction, arguing that these two processes also have different discourse functions: destressing is an anaphoric process which works independently of the focus set, that is, anaphoric elements are destressed irrespective of whether they are or are not in the focus set, whereas stress strengthening derives foci which are not already in the set. Neeleman and Reinhart (1998) formulate this first process as follows:

(48) A DP is destressed if and only if it is D-linked to an accessible discourse entity
(Neeleman and Reinhart 1998:338, ex. 364)

If an object needs to be destressed, it is incompatible with the most deeply embedded position because there, it would be stressed. Different languages have different ways of dealing with this problem. Dutch, Neeleman and Reinhart argue, makes use of the different word order options at its disposal. Thus, whenever there is an adverb present in a Dutch sentence, a D-linked object appears to the left of that adverb so that it is not assigned main sentence stress. Instead of scrambling, English moves the stress from the most deeply embedded constituent, the object, as in (49)-a, to another constituent, for example the verb, as in (49)-b. When the D-linked object is assigned sentential stress, as in (49)-a, the sentence is pragmatically infelicitous (as indicated with #).

(49) Context: There are beans growing in Ellie's garden

a. #She will harvest [the BEANS] today
Focus set: {#Object, VP, IP}

b. She will HARVEST [the beans] today
Focus set: {Verb, VP, IP}

According to Reinhart, then, Dutch scrambling can be seen as an equivalent of English contrastive stress and the interpretive effects which it involves a consequence of the focus considerations outlined above.

The reason why Dutch uses scrambling rather than stress-shift, even though the latter is a possibility in this language, is because stress-shift is considered to be marked in that it is a special, additional operation which applies once a sentence has been assigned stress already. In cases where stress-shift might apply, for example, when the verb is the intended focus or the object cannot be in focus (because, for example, it is D-linked), Dutch makes use of an option which the computational system makes available, namely scrambling. English does not have this option and is therefore forced to apply stress-shift. Thus, less economical operations, such as stress-shifting, only take place to satisfy an interface need and when a language has an alternative which can avoid the use of such an operation, this is preferred.

Reinhart (1995) remains agnostic as to whether scrambling should be seen as the result of movement or base-generation, but Neeleman and Reinhart (1998) argue for a base-generation approach (following Neeleman 1994; Neeleman and Weerman 1996, later published in 1999, see §2.3.1.3 for further discussion). Given that familiarity with Neeleman and Weerman's analysis will be necessary for the discussion of Neeleman and Weerman's (1997) study on the acquisition of scrambling in Chapter 3, I will briefly outline it here. On this analysis, whether a language has scrambling depends on its setting of the OV/VO parameter. This parameter specifies the direction in which abstract Case is checked, namely to the left or to the right, and it is accompanied by a universal strategy which determines the domain in which Case is checked, that is, whether this is in the syntactic phrase or in the phonological phrase. These two properties, Neeleman and Weerman argue, are crucially linked: the direction of checking implies the domain in which checking will take place. Hence, checking to the left implies a larger (syntactic) domain, with the consequence that such languages, that is, OV languages, allow objects to appear in a position which is not strictly adjacent to the verb (that is, they allow scrambling). VO languages, on the other hand, require the object to be directly adjacent to the verb, disallowing scrambling.[16] Neeleman and Reinhart exploit

[16] This implication follows from the fact that Case is checked at PF. Where possible, Case-checking takes place in a prosodic domain, but as a consequence of the way in which prosodic

the claim that both scrambled and non-scrambled word orders are freely generated by the OV setting of Dutch to support their argument that scrambling is more economical than stress-shifting in this language. Neeleman and Weerman's base-generation approach to scrambling will not be adopted here, however, for the following reason.[17] Like any analysis which links the availability of scrambling to OV word order (see also Haider and Rosengren 2003), Neeleman and Weerman's approach faces an empirical challenge in the form of the Slavic languages, such as Russian and Polish, which are VO but which also allow scrambling (Bailyn 2002). How these languages are analysed on such an approach would seem particularly important in an attempt to track the acquisition of an OV/scrambling language by speakers of a VO/no-scrambling language, as this would allow us to determine whether, for example, learners pass through a VO/scrambling stage.

For Reinhart and for Neeleman and Reinhart, there is a strong link between D-linkedness and scrambling. The existence of such a strong link has, however, been challenged in the recent literature (de Hoop 2000; 2003; Ruys 2001, see also the discussion of Zwart in §2.3.1.1.). De Hoop (2000) gives examples of definite NP objects which can optionally scramble on both an anaphoric (50) and a non-anaphoric (51) interpretation.

(50) Paul heeft een kat die de laatste tijd een gespannen indruk maakt
Paul has a cat who the last time a tense impression makes
'Paul has a cat who has been looking a bit tense recently.'

a. Misschien komt dat omdat hij zelden [de kat] aait
perhaps comes that because he rarely the cat strokes
'Perhaps that's because he rarely strokes the cat.'

b. Misschien komt dat omdat hij [de kat] zelden aait
perhaps comes that because he the cat rarely strokes

(51) Paul maakt de laatste tijd een gespannen indruk
Paul makes the last time a tense impression
'Paul has been looking a bit tense recently.'

a. Misschien komt dat omdat hij zelden [de kat] aait
perhaps comes that because he rarely the cat strokes
'Perhaps that's because he rarely strokes the cat.'

b. Misschien komt dat omdat hij [de kat] zelden aait
perhaps comes that because he the cat rarely strokes

(de Hoop 2000:157 ex. 7 and 8)

---

domains are formed, this is not possible when the object appears to the left of the verb. For more details, the reader is referred to Neeleman and Weerman's (1999) monograph.

[17] This contrasts with previous work (Unsworth 2003a), where (some of) the production data reported on in Chapter 5 were analysed using this approach.

In both of these examples, it is the IP which is in focus, not the verb. Recall that on Neeleman and Reinhart's analysis, the only difference between non-scrambled and scrambled word orders is that in the former, the object can be in focus and the verb cannot (cf. (45)-a), whereas in the latter, the verb can be in the focus and the object cannot (cf. (45)-b). In addition to the object and verb, respectively, the non-scrambled and scrambled order, also both allow the VP and IP to be in focus. Thus, Neeleman and Reinhart (1998:343) write:

> we would only expect a preference for scrambling when the verb needs to be contrastive (the only focus), or a preference for nonscrambling when the object needs to be the sole focus.

Hence, the examples in (50) and (51) would not constitute counter-evidence to Neeleman and Reinhart's (1998) analysis.[18] De Hoop (2000) does, however, provide two clear-cut counterexamples to Neeleman and Reinhart's account. In (52) the object is D-linked and the verb is contrastive, yet both orders are possible.

(52) Hoe gaat het met de review van Jans boek ?
how goes it with the review of Jan's book
'How's the review of Jan's book going?'

a. Nou, ik heb eindelijk [het boek] gelezen, maar
now I have finally the book read but

ik snap er nog niet veel van
I get still not much from
'Now, I've finally read the book, but I don't understand much of it.'

b. Nou, ik heb [het boek] eindelijk gelezen, maar
now I have the book finally read

ik snap er nog niet veel van
I get still not much from

(adapted slightly from de Hoop 2000:159, ex. 112)

[18] De Hoop suggests that (51)-b is also problematic because the non-D-linked object appears in scrambled position, where it does not receive stress, but according to Neeleman and Reinhart (1998:338, ex. 364) 'a DP is destressed if and only if it is D-linked to an accessible discourse entity'. It is not clear, however, that the NP in (51)-b is *not* linked to the preceding discourse, at least if we understand discourse in a fairly broad sense. It would seem that both sentences in (51) are likely to be uttered in contexts where Paul is known to have a cat. If this were not the case, for example if the speaker wanted to relay the same message as in (51) but was not sure whether the hearer knew that Paul had a cat, she would probably use an NP with a possessive pronoun, that is, *zijn kat* 'his cat', rather than the definite NP, *de kat* 'the cat'. In other words, the NP object may be 'D-linked' in the sense that it is part of long-term shared hearer and speaker knowledge.

In (53), the object is not D-linked and it is the sole focus. Although – as predicted by Neeleman and Reinhart – there may be a preference for (53)-a over (53)-b, both orders are perfectly acceptable.

(53) Heeft je buurman gisteren de deur geverfd?
has your neighbour yesterday the door painted
'Did your neighbour paint the door yesterday?'

a. Nee, maar hij heeft gisteren wel [de dakgoot] geverfd
no but he has yesterday indeed the gutter painted
'No, but he did paint the gutter yesterday.'

b. Nee, maar hij heeft [de dakgoot] gisteren wel geverfd
no but he has the gutter yesterday indeed painted

(adapted slightly from de Hoop 2000:159, ex.113)

These two examples constitute a problem for Neeleman and Reinhart's analysis.

Thus, although it is clear that discourse-relatedness does have an effect on which word order (scrambled vs. non-scrambled) is preferred, it is certainly not as deterministic as Neeleman and Reinhart's (1998) account suggests. Note, however, that, it is not the case that Neeleman and Reinhart claim that scrambling is 'triggered' by discourse-relatedness (cf. Schaeffer 2000). They argue that the grammar allows both scrambled and non-scrambled word orders and that these are used at the PF interface such that particular foci can be expressed without having to invoke any costly stress-shifting operations. Note that this approach quite nicely accounts for the difference between scrambled and non-scrambled definite objects in negative sentences. Recall that when definite objects are scrambled across negation, as in (54)-a, the interpretation is one of sentential negation, whereas when the definite object is not scrambled, as in (54)-b, the sentence is interpreted as expressing contrastive negation (cf. §2.1.1).

(54) a. Bob heeft [het onkruid] niet WEGgegooid
Bob has the weeds not away-thrown
'Bob didn't throw away the weeds.'

b. Bob heeft niet [het ONKRUID] weggegooid
Bob has not the weeds away-thrown
'Bob didn't throw away the WEEDS.'

On the assumption that negation associates with focus, the difference between (54)-a and (54)-b could thus be analysed as follows. The non-scrambled form is preferred when the object needs to be the sole focus. When this is not the case, the scrambled order is used.

2.3.3 *Semantic approaches to scrambling*

This section focuses on two analyses of scrambling which argue that scrambled and non-scrambled indefinite objects are semantically different from each other. De Hoop (1992) argues that the semantic type of an object follows from the type of Case which that object is assigned and this determines whether an object can scramble or not. Van Geenhoven (1998) claims that scrambling serves to indicate which semantic type a given object is of. The semantic properties of scrambled and non-scrambled indefinites are central in this thesis and for this reason, I will examine these two proposals in some detail.

2.3.3.1 De Hoop (1992)

De Hoop's (1992) approach combines insights from generative syntax (relating to Case) and model-theoretic semantics (type theory).[19] Rather than viewing the relationship between syntax and semantics as a strict one-to-one mapping (cf. Diesing 1992), de Hoop maintains that although there is a strong correlation between the two components, there are parts of syntax and semantics which are independent of each other.[20] She observes that when weak NP objects scramble,[21] they have one of four different readings – referential, partitive, generic and generic collective – all of which can be categorised as 'strong':

(55)	a.	…dat	de politie	[een kraker]	gisteren	opgepakt	heeft	
		…that	the police	a squatter	yesterday	up-picked	has	
		'*…that the police picked up a squatter yesterday.*'						
								(referential)

---

[19] Expanding on her earlier research, discussed in this section, de Hoop (2000; 2003) has recently adopted an Optimality Theory (OT) approach to scrambling. This will be discussed in more detail in Chapter 7 (§7.4.2) in the context of de Hoop and Krämer's (to appear) recent paper on the acquisition of scrambling from an OT perspective.

[20] Diesing's (1992; 1997) 'Mapping Hypothesis' adopts a very strict one-to-one mapping from syntax to semantics. This approach to scrambling will not be discussed here. For problems with this proposal and reasons why it cannot account for the Dutch scrambling data, see de Hoop (1992), van der Does and de Hoop (1998), Reinhart (1995) and Ruys (2001).

[21] Weak NPs are those with determiners which are allowed in existential sentences, as in (i), and strong NPs are those with determiners which are not allowed in such sentences, as in (ii) (Milsark 1977).

(i) There are some / three / no birds in the garden
(ii) *There are all / most / both birds in the garden

Weak NPs are said to have a strong reading when, in certain contexts, they behave like strong NPs.

b. …dat de politie [twee krakers] gisteren opgepakt heeft
…that the police two squatters yesterday up-picked has
'…that the police picked up two squatters yesterday.'

(partitive)

c. …dat de politie [krakers] altijd oppakt
…that the police squatters always up-picks
'…that the police always pick up squatters.'

(generic)

d. …dat de politie [tien krakers] altijd oppakt
…that the police ten squatters always up-picks
'… that the police always pick up ten squatters.'

(generic collective)

(de Hoop 1992:50, ex. 143-146)

De Hoop's (1992) analysis works as follows: NPs are assigned either 'strong' or 'weak' Case, where strong Case should be understood as structural Case licensed at S-structure and weak Case as a default structural Case licensed at D-structure in particular syntactic configurations (for example, when the object is directly adjacent to the verb). Object NPs marked with strong Case are allowed to scramble (but crucially, they do not have to) and they receive a strong (quantificational) reading. This, de Hoop (1992) claims, is because strong Case serves as an obligatory 'type-lifting' operator.

Following Partee (1987), de Hoop assumes that different NPs can be of different denotational types, of which there are three basic ones: *e* ('referential'), <*e*,*t*> ('predicative') and <<*e*,*t*>, *t*> ('quantificational').[22] These types are related to each other by means of general 'type-shifting operations', which serve to map from one meaning to another for one and the same NP. Definite NPs are generally considered to be of type *e*, whereas indefinite NPs are, according to de Hoop (1992), of type <<*e*,*t*>,<*e*,*t*>>, that is, predicate modifiers.[23] The basic idea is that when an NP object bears strong Case, it is always 'lifted' to a (higher) type, namely the quantificational type, <<*e*,*t*>, *t*>. Furthermore, it can appear in both scrambled and non-scrambled positions, because strong Case is licensed at S-structure. Any object NP bearing weak Case, on the other hand, is forced to remain in its base (D-structure) position because it is here that weak Case is assigned. Thus, on de Hoop's (1992) account, when indefinite NP objects have a specific reading, they must have

[22] Readers who are not familiar with type theory might what to think of NPs being of different denotational types in the same way that different combinations of elements in a sentence can be different types of phrases.

[23] In more recent work (e.g. van der Does and de Hoop 1998), de Hoop analyses non-scrambled indefinites as type <*e*,*t*>, which are incorporated into the verb along the lines of Van Geenhoven (1998); see below for relevant discussion.

strong Case, because this automatically lifts them from their usual (predicate modifier) type of $<<e,t>,<e,t>>$ to the generalised quantifier type, $<<e,t>, t>$. Because they have strong Case, which is licensed at S-structure, these objects are allowed to – but crucially do not necessarily have to – scramble.

Ruys (2001) argues against de Hoop's claim that scrambled weak NPs must be strong. First, he provides examples of weak indefinite NP objects which remain weak when scrambled. In (56), the scrambled indefinite object *een of andere ziekte* 'some disease or other' has a weak interpretation. It cannot be generic because this is incompatible with the determiner *een of ander*. Neither can it be partitive because the sentence does not preclude that doctors treat all (rather than some) diseases with penicillin. The only strong reading left (on de Hoop's account), a specific reading, is also not possible (in the classic (widest scope) sense) because the object is interpreted within the scope of the structurally higher quantified subject, *elke arts* 'every doctor' (see below for discussion of similar examples).

(56)	dat	elke arts	wel	[een of andere	ziekte]
	that	every MD	(particle)	one or other	disease

	meestal	met pencilline	behandelde
	usually	with penicillin	treated

'*…that every MD treated s(o)m(e) disease or other usually with penicillin.*'

(Ruys 2001:51, ex. 10)

Secondly, Ruys shows that there is also a shift in meaning when strong indefinite NP objects scramble, suggesting that scrambling affects the meaning of a structure as a whole, rather than the meaning of an NP alone:

(57)	a.	…dat	Jan	drie keer	[bijna	alle glazen]	omgooide
		…that	John	three times	almost	all glasses	over-knocked

'*…that on three occasions, John knocked over almost every glass.*'

	b.	…dat	Jan	[bijna	alle glazen]	drie keer	omgooide
		…that	John	almost	all glasses	three times	over-knocked

'*…that almost every glass, John knocked over three times.*'

In other words, the interpretive effects found for scrambled indefinite NP objects are at least partly a consequence of the relative scope of the object NP and any scope-taking elements in a given sentence. Thus, as Ruys acknowledges, the search for the explanation behind these effects is a multi-faceted enterprise.

2.3.3.2 Van Geenhoven (1998), Krämer (2000)

Van Geenhoven's approach differs from that of de Hoop in that the link between syntax and semantics is much weaker. This is the analysis which is adopted (and expanded upon) by Krämer in her study of the L1 acquisition of the comprehension of scrambled indefinite objects and its basic insights will also be adopted here. The fundamental idea underlying Van Geenhoven's approach is that unless there is some indication to the contrary, indefinite NP objects are always interpreted as being part of the verb (that is, as predicative indefinites) and that scrambling is one such way of indicating that this is not the case, that is, that the indefinite NP object has a different interpretation, namely that of a free variable.

Following Kamp (1981) and Heim (1982), Van Geenhoven assumes that indefinites are variables, but unlike these authors, she assumes that they do not introduce variables themselves. Working in a Discourse Representation Theory framework (Kamp and Reyle 1993), Van Geenhoven assumes that indefinite NPs get their interpretation from the context in which they appear: the indefinite's variable is introduced either by the semantically-incorporating verb or by 'accommodation'. On Van Geenhoven's account, each transitive verb has two lexical entries: one where it is semantically-incorporating and one where it is not, that is, one which is truly transitive. When an indefinite direct object is incorporated by the verb, the property expressed by the indefinite becomes a condition on the argument slot of the verb, that is, it restricts the values of the argument which the verb takes. For example, if the indefinite *a ball* incorporates into the verb *throw*, as in (58), the result is the predicate *ball-throw*, and so the indefinite can no longer be regarded as a true argument.

(58) $\lambda P \lambda x. \exists y\ [P(y) \wedge \text{throw}\ (y)(x)]$

Consequently, non-scrambled indefinites, which are incorporated, cannot escape the scope of negation or a structurally higher adverbial precisely because they are incorporated into the verb and the verb is in the scope of negation or of the adverb. Free variable indefinites, on the other hand, are not incorporated. They are true arguments of the verb and their variable is bound at some higher level. For this purpose, Van Geenhoven (1998) adapts van der Sandt's (1992) theory of accommodation. If a free variable is not bound, it is uninterpretable at LF. On this account, then, it is the process of accommodation which is crucial to the interpretation of scrambled indefinites.

Van der Sandt (1992) uses accommodation as a repair mechanism to resolve the presuppositions triggered by definite NPs. For example, the definite NP in (59), *Jack's children* presupposes that Jack has children.

(59) If Jack is bald, then all of Jack's children are bald

(Van Geenhoven 1998:199, ex. 112)

The mechanism of accommodation introduces this discourse referent to the context, such that the anaphoric expression *Jack's children* can be identified, as in (60). This addition is generally assumed to operate at the top-most (that is, a global) level; in other words, this addition is placed at the highest point in the discourse representation.

(60) [Jack has children x]$_{\text{accommodation}}$ If Jack is bald, then all of x are bald
(Van Geenhoven 1998:199, ex. 113)

Van Geenhoven applies accommodation to indefinites. She claims that when an indefinite object is not incorporated into the verb, for example, when it is scrambled, as in (61)-a, its variable is introduced by accommodation, illustrated in (61)-b.

(61) a. Jos zou [een grote appelboom] morgen moeten gaan planten
Jos would a big apple-tree tomorrow must go plant
'Jos should plant a (certain) big apple-tree tomorrow.'

b. [Er is een grote appelboom x]$_{\text{accommodation}}$
there is a big apple-tree

Jos zou x morgen moeten gaan planten
Jos would tomorrow must go plant
'There is a big apple-tree x. Jos should plant x tomorrow.'

The existential reading of the object in (61)-b is due to the assumption that variables introduced at the top-most level are interpreted existentially (cf. Kamp and Reyle 1993). The accommodation analysis thus straightforwardly accounts for why scrambled indefinites, that is, free variable indefinites, escape the scope of any sentential adverbials or negation. On Van Geenhoven's account, scrambling is therefore simply a language-specific means of marking the predicate-free variable distinction.[24] In (62), the non-scrambled indefinite has been semantically incorporated into the verb.

(62) Jos gaat morgen [een grote appelboom] planten
Jos goes tomorrow a big apple-tree plant
'Jos is going to plant a(ny) big apple-tree tomorrow.'

On Van Geenhoven's approach, the object must be directly adjacent to the verb for incorporation to take place. The verb introduces and binds the object's variable. The verb is in the scope of the adverbial and consequently, the incorporated object is, too. This contrasts with (61), where the indefinite object is clearly outside of the scope of the adverbial, indicating that it must be a free variable indefinite, that is, its variable is introduced by accommodation and existentially interpreted at a higher level.

---

[24] Others include overt objective Case marking (e.g. Turkish, Finnish) as discussed by de Hoop (1992) and the overt incorporation found in West Greenlandic, documented by Van Geenhoven (1998) herself. See §2.5.2 for a discussion of how free variables are marked in English.

Krämer expands Van Geenhoven's analysis to account for the following observations. When the context contains a set of objects to which the scrambled indefinite can be related, it is almost impossible *not* to relate the scrambled indefinite to that set. This is what Krämer refers to as a 'source-set' guise. When no such set is present, a 'specific' guise is understood, that is, the object 'refers to an individual that the speaker can identify' (based on Fodor and Sag 1982; Krämer 2000:8). This is illustrated in (63)-a and (63)-b for the 'source-set' and 'specific' guises, respectively.

(63) a.

In	haar	tuin	heeft	Eloise	een	heel	bos	staan.
in	her	garden	has	Eloise	a	whole	forest	stand

Ze	vindt	het	leuk,	maar	ze	gaat	[een boom]	morgen
she	finds	it	nice	but	she	goes	a tree	tomorrow

eruithalen	(want	hij	is	ziek)
it-out-take	because	he	is	diseased

'Eloise has a whole forest in her garden. She thinks it's nice, but she's going to remove a tree from it tomorrow.(because it's rotten).'

b.

Eloise	gaat	[een	boom]	morgen	uit	de	tuin	halen
Eloise	goes	a	tree	tomorrow	out	the	garden	take

'Eloise is going to take a tree out of the garden tomorrow.'

Krämer argues that the difference between the 'source-set' and 'specific' guises are the consequence of what she dubs the 'Context Requirement', given in (64), and that this in turn follows from the need for 'bridging' (Clark 1977).

(64) Context Requirement
A high indefinite NP is only felicitous if the context provides some element to which it can be related.
(Krämer 2000:39, ex. 60)

Following Schouten (1998), who studied indefinite objects in scrambled or subject position in written Dutch, Krämer claims that the link between the scrambled indefinite object NP and the context is conceptual, rather than grammatical. Specifically, she claims that 'the source of the Context Requirement lies in the fact that the context must provide input for bridging' (Krämer 2000:43). The inferential process of 'bridging', proposed by Clark (1977), refers to how a hearer identifies an antecedent for a pronoun or definite NP by making a series of inferences based on the information which has been provided by the speaker. It is assumed that the speaker will provide the necessary information to make this inferential process possible. This is captured in the 'Given-New Contract':

(65) The speaker agrees to try to construct the Given and New information of each utterance in context so that the listener is able to compute from memory the unique Antecedent that was intended for the Given information […] .

(Clark 1977:413)

Thus, in (66), the hearer must infer that the speaker expects her to find the antecedent for *the shed* and that the speaker has provided enough information to make this possible. In this case, the hearer may conclude that *the shed* to which the speaker is referring must be on the allotment which the speaker has mentioned, because, for example, this is where sheds are often located. In other words, the definite NP *the shed* is accommodated by means of bridging.

(66) Melinda went down to the allotment. The shed was on fire.

Krämer proposes that a similar process occurs with free variable indefinites. In this case, bridging serves to identify the accommodation site of the free variable indefinite.[25] This is directly observable in cases where free variable indefinites are accommodated to some level other than the top (global) level discussed above, that is, in cases where intermediate readings are available (Abusch 1994; Ruys 1992). Van Geenhoven observes that in addition to a wide and narrow scope reading, the object *een boek* 'a book' in (67) can also have an intermediate reading.

(67)

Iedere auteur	in deze kamer	veracht	iedere uitgever	die	[een boek]
every author	in this room	despises	every publisher	who	a book

dat	als	pornografisch	was	bestempeld	niet	wilde	uitgeven
that	as	pornographic	was	deemed	not	wanted	publish

Narrow scope:	For each publisher p for whom there is a book that was deemed pornographic, it is the case that each author in this room despises p.
Intermediate scope:	For each author a in this room, there is a book b that was deemed pornographic such that a despised every publisher who would not publish b.
Wide scope:	There is a book b that was deemed pornographic, and each author in this room despises every publisher who would not publish b.

According to Van Geenhoven, this is a consequence of the conceptual link which exists between *een boek* 'a book' and *iedere auteur* 'every author'. This conceptual link provides a covert anaphoric element which is bound by the quantified subject *iedere auteur* 'every author' and which traps the free variable indefinite (so-called 'implicit trapping' (Sandt 1992)). The precise details of the analysis need not concern us here. What is important to note is that in

[25] How exactly the accommodation site is defined is left open by Krämer.

sentences such as (67), the free variable indefinite can be accommodated globally (as discussed earlier in this section) or at this intermediate level (as a result of 'implicit trapping').

Krämer notes that without any supporting context, all three readings are possible, but when this sentence is embedded in a context such as (68), the wide scope reading is much more likely. This is because the object, *een boek*, is linked to 'the work of the famous author X' (which presumably includes a set of books) as a result of bridging.

(68) I will tell you why the atmosphere is so tense: There has been a heated discussion in the media about the literary qualities of the work of the famous author X.

Whereas accommodation is driven by grammatical requirements, that is, as a repair mechanism which adds information to the discourse so that elements which would otherwise be uninterpretable can be interpreted, bridging is driven by communicative requirements, namely a desire on the part of the hearer to reconstruct the speaker's intended message. Bridging occurs more or less automatically, as illustrated in (69), where the anaphoric elements *it* and *the chandeliers* are each related to the antecedent available in the context, *a little boat*, even though this may be rather strange (as in (69)-a) or downright impossible (as in (69)-b).

(69) a. I was rowing [a little boat]. It said to me …
b. I was rowing [a little boat]. #The chandeliers sparkled brightly.
(Krämer 2000:45, ex. 71)

In the case of scrambled indefinites, the automaticity of bridging explains why, if there is a set in the context to which the scrambled indefinite can be related, it will be. This is what occurs when the sentence in (67) is preceded by the context in (68). Also, in (63)-a above, the scrambled indefinite object NP, *een boom* 'a tree', is automatically assumed to be part of the forest mentioned in the first sentence. If there is no such set in the context, then assuming the Given-New Contract, the scrambled indefinite will be interpreted by default as 'specific', because the hearer will assume that the speaker considers the information which she has provided to be sufficient. If there is no object in the immediate physical context which the speaker could be referring to, the hearer will assume that the speaker must have a specific object in mind. Thus, in (63)-b above, it is assumed that the speaker knows that there is a particular tree which Eloise will remove. If she did not, the speaker would have provided enough information in the context for this to be clear. It is in this way, Krämer claims, that bridging leads to the Context Requirement (64). Note, also, that the marginal acceptability for some native speakers of contextless utterances containing scrambled indefinite object NPs may also be explained by bridging and the Context Requirement. Some native speakers may simply be better at coming up with contexts which render utterances containing scrambled indefinite objects felicitous, whereas others are not (Krämer 2000:50)

The combination of Van Geenhoven (1998) and Krämer (2000) provides a semantic-pragmatic analysis of the interpretive properties of scrambled indefinite object NPs in Dutch. On this analysis, scrambling is a language-specific means of marking a particular reading, namely a free variable reading, of an indefinite object NP. Via the notion of bridging, it is possible to account for the different interpretive guises which such indefinites have. Note, however, as Krämer herself does (2000:48, fn. 18), that her application of bridging in this domain could, albeit in very different terms, also be combined with a different analysis of indefinites, such as the quantificational account put forward by de Hoop (1992). For example, it could serve the purpose of identifying the scrambled indefinite's domain restriction.

The semantic approaches reviewed here seem to be able to capture most (if not all) of the properties of scrambled indefinite object NPs. The choice between one over the other is partly contingent on how the relation between syntax and semantics should be viewed (see Krämer 2000:36ff. for relevant discussion). Although de Hoop rejects a one-to-one mapping between syntax and semantics (cf. Diesing 1992; see fn. 20), there is nevertheless a clear link between the two: scrambled indefinites always receive a strong (quantificational) reading. For Van Geenhoven, on the other hand, scrambling is simply a language-specific means of marking an indefinite as having a free variable rather than a predicative interpretation. Also crucial in distinguishing between these two approaches is the importance which should be attributed to intermediate readings. Van Geenhoven's analysis can account for these, whereas de Hoop's analysis cannot.

2.4 Approach adopted here

In this section, I outline the analysis which will be adopted here. It is clear from the preceding discussion that there is no one single property which is shared by all scrambled objects and therefore, it cannot be claimed that scrambling is *driven* by anything (de Hoop 2000; 2003). I will therefore assume that the scrambled and non-scrambled orders are simply generated by the grammar and that these are exploited in the following two ways. First, following Van Geenhoven (1998), scrambling of indefinite objects is assumed to disambiguate meaning, that is, to indicate that an indefinite object is of a certain semantic type, namely a free variable on Van Geenhoven's analysis or a quantificational $<<e,t>, t>$ type on de Hoop's approach. Given that scrambled and non-scrambled definites mean the same truth-conditionally (van der Does and de Hoop 1998), the scrambling of definite NP objects across negation, which results in sentential negation rather than contrastive negation, cannot mark a difference in the semantic type of the object. Rather, the definite object NP scrambles as a result of the interaction between negation and focus.

Therefore, the second way in which I will assume scrambling is exploited is, in the spirit of Neeleman and Reinhart (1998) and Reinhart (1995), to derive a particular focus. Specifically, the scrambled order (with neutral stress) has a sentential rather than contrastive negation interpretation for the following reason. For the object to have a contrastive reading, negation must be associated with the object only. On the assumption that negation associates with focus, this would require only the object to be in focus and this is only possible when it occurs in non-scrambled position. When the object is in the scrambled position, it cannot be the only focus, and therefore, a contrastive negation reading is not possible.

What about the syntax of scrambling? In the preceding paragraph, I assumed that the scrambled and non-scrambled orders are generated by the grammar without stating how or why. There are two possibilities, either movement or base-generation. In the discussion in §2.3.2.2, I suggested that base-generation approaches which link the option of scrambling to OV word order, such as Neeleman and Weerman (1999), are problematic, especially in the context of the present study, because they cannot account for, or at least do not provide any (explicit) analysis of, VO scrambling languages. Although this does not preclude the possibility that such an analysis could be developed or that scrambled objects may be base-generated for some other reason, in this thesis, I will assume that scrambled orders are the result of movement.[26, 27] A movement approach to scrambling could (but need not necessarily) be implemented in a minimalist, feature-driven approach in terms of the optional insertion of a semantically vacuous [scrambling] feature which is mandated by the type of LF and PF interface requirements outlined above (as in Hopp (2003), following Grewendorf and Sabel (1999) and Haider and Rosengren (1998)). On this type of analysis, the object moves to some VP-external position to check its strong [scrambling] feature ([Spec, AgrOP] on Hopp's approach, following Collins and Thráinsson (1996), amongst others).[28]

To summarise: In this thesis, scrambling is assumed to be a movement operation which serves to indicate (i) that an indefinite object is of a particular semantic type, that is, it is not incorporated into the verb, and (ii) in the case of

---

[26] This is supported by recent evidence from the processing literature (see e.g. Miyamoto and Takahashi 2002).

[27] In the discussion of Neeleman and Weerman (1997) in Chapter 3, I show that the L1 acquisition data do not conform to the predictions made by Neeleman and Weerman's analysis about the link between scrambling and OV.

[28] As Haider and Rosengren (2003) point out, feature-driven approaches to scrambling which involve the optional insertion of a [scr(ambling)] feature to trigger scrambling beg the question (e.g. Sauerland (1999)). Furthermore, they are often too strong, because they incorrectly predict obligatory scrambling. On approaches such as that of Hopp (2003), this problem is obviated because the insertion of the scrambling feature is mediated by interface requirements and the interplay between these requirements is mediated by interface economy (in the spirit of Adger (1996) and Reinhart (1995) amongst many others).

definite objects in negative sentences, that the definite object is not in focus and will not be interpreted in the scope of negation.

The final section of this chapter spells out the differences between Dutch (the target language for the (L1 and L2) learners in this thesis) and English (the L1 for the L2ers).

2.5 Differences between Dutch and English

2.5.1 Syntactic structure

Following traditional analyses of Dutch (e.g. Koster 1975), I will assume that Dutch is underlyingly head-final (OV), whereas English is head-initial (VO). English does not have scrambling, that is, it does not have an overt movement operation which induces that same type of interpretive effects as scrambling in Dutch. Adverbs are assumed to be VP-adjoined in both English and Dutch, and following Pollock (1989) and Ouhalla (1990), negation is assumed to head its own phrase, NegP, which is also VP-adjoined. Recall that the English equivalents of scrambled (and non-scrambled) sentences generally involve a sentence-final adverb, as in:

(70) a. Cécile planted [a rose] yesterday
b. Simon watered [a rose] twice / two times

As noted by Haider (2004), the exact position of post-verbal adverbials is very unclear. Here, it will be assumed that they are right-adjoined to the VP.[29]

[29] How to analyse post-verbal adverbials is a perennial problem, one which poses a particular challenge to any approach adopting Kayne's (1994) anti-symmetry approach, such as Zwart (1997), for example. Post-verbal adverbials are often analysed as being left-adjoined to VP with the verb raising to little *v* and the object to [Spec, AgrOP] to check its Case features (e.g. Bobaljik 1995). Such an approach is problematic for two reasons. First, it requires an additional explanation for verbs which obligatorily select a PP, as in (i), which do not have Case to check (B.D. Schwartz, p.c., 23 March 2005).

(i) Rick resided in Edinburgh last year

Second, it implies that English and Dutch are essentially the same in that both allow overt movement to [Spec, AgrOP], which means that in both languages, the optional insertion of adverbials to the right and to the left of this position should also possible. If this were the case, the explanation for why interpretive effects are observed in Dutch but not in English would have to be sought elsewhere. One option would be to account for this difference in the properties of adverbs and of intonation and focus, as Zwart (1997) does, but this approach was found to be unsatisfactory in this regard. It would seem that if English and Dutch both have overt movement to [Spec, AgrOP] for Case-checking purposes, Dutch must also allow some other movement operation, in which case it must be this which is associated with the relevant interpretive effects (E. Ruys, p.c., 10 June 2005). Note also that on the above scenario, where both English and Dutch have overt movement to [Spec, AgrOP] and the optional insertion of adverbials to the left and right of this projection, the syntax of scrambling should not cause any problems for English-speaking L2ers of Dutch, that is, it might be expected that, assuming

2.5.2 Interpretive effects

Scrambling in Dutch is used to indicate that an indefinite object is of a particular semantic type, which means that it has a specific reading. In English, the specific and non-specific readings are available for the same order. The sentence in (70)-b can thus refer to a particular flower Simon watered twice or to two different flowers. The first (specific) reading can be paraphrased as 'there is a flower such that Simon watered that flower on two occasions' and the second (non-specific) reading as 'Simon took part in two flower-watering events'. The non-specific reading involves a regular indefinite of type <*e*, *t*> incorporated into the verb. Because it is incorporated in the verb, it is in the scope of the higher, (right-adjoined) adverbial and is hence interpreted non-specifically. To derive the specific reading, several options are available. I will concentrate on two. The first would be to claim that there is no specific reading as such, but that what appears to be such a reading is actually a consequence of entailment relations. The second is to adopt Van Geenhoven's analysis for English. Each of these options is discussed in turn.

Recall that the indefinite object NP in the non-scrambled sentence in Dutch (71)-a is incorporated into the verb and is consequently in the scope of the adverbial *twee keer*. There are two situations in which this sentence is true, namely one involving two different roses and another involving one and the same rose. Thus, (71)-a refers to two flower-watering events and whether this involves two different roses or one and the same rose is irrelevant.

(71)	a.	Simon	heeft	twee keer	[een roos]	besproeid
		Simon	has	two times	a rose	watered
		'Simon watered a(ny) rose twice.'				
	b.	Simon	heeft	[een roos]	twee keer	besproeid
		Simon	has	a rose	two times	watered
		'Simon watered a (certain) rose twice.'				

This is different from (71)-b, where the indefinite object NP takes wide scope with respect to the sentential adverbial *twee keer*, that is, the only situation in which (71)-b is true is one where one and same rose is watered twice. In the English sentence in (70)-b, the object is also incorporated into the verb and it is possible that the availability of a verifying model in which one particular rose has been watered twice is due to the non-specific reading being compatible with such a model, rather than the specific reading being genuinely available (Reinhart 1976). This cannot be the case for the sentence in (72), however.

(72) Brigit did not pick a rose

---

transfer, these learners are targetlike from the earliest stages of acquisition. The data presented in Chapter 5 indicate that this prediction is not borne out, however.

In (72), it is not the case that models in which one particular rose is not picked generally verify the non-specific reading, that is, if there is one rose which is not picked, it is not necessarily the case that no roses were picked.[30] For (72), then, the availability of the specific reading cannot be because the situations which make the non-specific reading true also make the specific reading true. This suggests that a specific reading is genuinely available (see §2.2 for examples showing that this is the case for (70)-b). Experimental work reported on in Chapters 3, 6 and 7 supports this claim.

Assuming then that a specific reading is available, how is this derived? I will focus on Van Geenhoven's approach.[31] On this analysis, the specific and non-specific readings of the indefinite object in English would be analysed in the same way as the scrambled and non-scrambled sentences in Dutch, that is, on the specific reading, the indefinite is a free variable which is interpreted via accommodation, and on the non-specific reading, it is incorporated into the verb. Recall that on Van Geenhoven's analysis, indefinite objects are assumed to be incorporated into the verb unless there is some indication otherwise. For English, Van Geenhoven mentions two instances where there is some such indication. First, the determiner which is used may serve to indicate that a free variable interpretation is intended. Recall that the plural indefinite determiner *some* cannot take narrow scope with respect to negation. Thus, the object in (73) can only have a specific reading, that is, it can only mean that there are some flowers which Berit did not pick.

---

[30] The only situation where this would be the case is where there is only one rose in the context and that rose was not picked.

[31] Other ways of deriving the specific reading include assuming that indefinite NPs are lexically ambiguous along the lines of Fodor and Sag (1982) or that covert movement such as Quantifier Raising takes place (May 1977) (but see Hornstein 1995; Reinhart 1997). On the former analysis, the indefinite object in (70)-b, *a rose*, would be interpreted as any entity which has the property of being a rose on the non-specific reading, and on the specific reading, it would be interpreted in the same way as a demonstrative such as *this rose*. This latter interpretation ('referential' in Fodor and Sag's terms) involves 'an unscoped element which does not bind variables and is insensitive to island boundaries, but which happens to entail a maximally wide scope quantifier reading' (Fodor and Sag 1982:379). The denotation of a referential indefinite is:

> the individual the speaker 'has in mind', as long as the descriptive content is true of that individual. Thus a referential indefinite can be regarded as an indexical phrase, where the function that determines its referent depends both on its descriptive content and on its context of use, and the latter is taken to include the psychological state of the speaker […]. (Fodor and Sag 1982:284)

As Krämer (2000: 7-8) notes, this characterisation successfully captures most of the available readings for scrambled indefinite NP objects in Dutch, but it cannot, however, account for the availability of intermediate scope readings (cf. §2.3.3.2).

(73) Berit didn't pick some flowers

Also, as noted in §2.2, the use of an adjective such as *certain* or *particular* or for some speakers, an overt partitive, can also serve to indicate that a specific (free variable) reading is intended.

(74) a. Cécile planted [a certain rose] yesterday
b. Simon watered [a certain rose] twice
c. Brigit didn't pick [a certain rose]

(75) a. Cécile planted [one of the roses] yesterday
b. Simon watered [one of the roses] twice
c. Brigit didn't pick [one of the roses]

Using the cardinal/stressed indefinite *one*, as in (76), might induce a specific reading if the hearer has some reason to believe that the speaker has a particular object in mind. When used in a sentence containing negation, however, it can also be understood as expressing constituent negation. The sentence in (76)-c would therefore be paraphrased as 'the number of roses which Brigit picked is not one'. This is similar to the Dutch equivalent, given in (77).

(76) a. Cécile planted [one apple-tree] yesterday
b. Simon watered [one flower] twice
c. Brigit didn't pick [one rose]

(77) Brigit heeft niet [één roos] geplukt
Brigit has not one rose picked
'Brigit didn't pick one rose.'

It is also possible that the discourse provides some indication of the intended free variable interpretation:

(78) John ate everything his mother gave for lunch but he didn't eat an apple and a pear
(Van Geenhoven 1998:208, ex. 232, attributed to Dayal (p.c.))

Accounting for the availability of both readings in English on de Hoop's (1992) analysis would presumably work as follows: the indefinite object can be assigned either weak or strong Case and this will determine which reading is available.

Concerning definite objects and scrambling across negation, Dutch offers different word order possibilities which can be exploited in order to convey a sentential rather than a contrastive negation interpretation. In (79)-a, where the definite object has scrambled across negation, the interpretation is one of sentential negation, whereas in (79)-b, the interpretation is one of contrastive negation. As noted in the preceding section, this is assumed to be the result of the interaction between focus and negation.

(79) a. Bob heeft [het onkruid] niet WEGgegooid
Bob has the weeds not away-thrown
'Bob didn't throw away the weeds.'

b. Bob heeft niet [het ONKRUID] weggegooid
Bob has not the weeds away-thrown
'Bob didn't throw away the WEEDS.'

English does not have the option of scrambling and consequently the difference between sentential and contrastive negation is achieved using stress-shift, as illustrated in (80). In (80)-a, the stress falls on the most deeply embedded constituent, the object, and the interpretation could either be one of sentential negation or one of contrastive negation. (To unambiguously express a contrast on the object, the object would have to be heavily stressed.) In (80)-b, where the stress has shifted from the object to the verb, the object can no longer be the sole focus of the sentence and on the assumption that negation associates with focus, a contrastive negation interpretation for the object is ruled out. This also holds for (80)-c, where the stress has shifted to the negator.

(80) a. Brigit didn't pick a ROSE
b. Brigit didn't PICK a rose
c. Brigit did NOT pick a rose

(80)-b is thus interpreted as expressing sentential negation or contrastive negation on the verb and (80)-c as expressing sentential negation. (Again, to unambiguously express a contrast on the verb in (80)-b, the verb would have to be heavily stressed.)

2.5.3 Summary

The differences between Dutch and English are as follows. Dutch has a movement operation (scrambling) which is associated with interpretive effects. These effects include the following. First, scrambled indefinite objects are interpreted as semantic objects of a particular type, namely as free variables or of type $<<e,t>, t>$ (quantificational); and second, definite objects appear in scrambled position in negative sentences so that a constituent negation interpretation is avoided. English does not have such a movement operation. The specific (or free variable) interpretation of indefinite objects in that language are indicated by other means, either by the use of a different determiner, the addition of an adjective such as *certain*, or by the context in which they are uttered. The difference between sentential and constituent negation in sentences with definite objects is achieved using stress-shift.

2.6 Summary

This chapter outlined the properties of direct object scrambling in Dutch which will be the focus of the experimental work discussed in the remainder of this thesis. Some of the current theoretical analyses of these properties were presented and evaluated. Trying to remain as analysis-neutral as possible, whilst still capturing the relevant range of data, I adopted the following approach. In Dutch, the grammar generates both scrambled and non-scrambled word orders by means of syntactic movement. These two orders are exploited for (at least) two purposes: either to indicate that an indefinite NP object is of a particular type, namely a free variable or a quantificational type, or to ensure that a definite object is not in focus and is out of the scope of negation, that is, to ensure that a sentential rather than constituent negation interpretation is intended. This contrasts with English, where such alternations in word order are not available. In that language, the different meanings associated with scrambled and non-scrambled orders in Dutch are expressed using other means, all of which are also available in Dutch but less preferred.

In the following chapter, I consider how these properties of Dutch are acquired.

CHAPTER 3

THE ACQUISITION OF SCRAMBLING

Introduction

This thesis compares child L2 acquisition with, on the one hand, adult L2 acquisition, and child L1 acquisition, on the other. It compares and contrasts these three groups in their acquisition of scrambling in Dutch. More specifically, it investigates the acquisition of the scrambling of direct objects across negation and the frequency adverbial *twee keer* 'twice', as measured in production and comprehension tasks. As detailed in the previous chapter, the option of placing a direct object in a scrambled position is assumed to be generated by the grammar as a result of movement to some VP-external position. All the cases of scrambling under consideration here have interpretive effects. When definite NP objects are scrambled across negation, the resulting meaning is generally one of sentential negation; scrambled indefinite NP objects are restricted to a specific interpretation. The interpretation of scrambled indefinite object NPs is affected by discourse/pragmatic considerations. Specifically, the interpretation assigned to scrambled indefinite object NPs is constrained by contextual information. The acquisition of scrambling thus represents a multi-faceted process which reflects the interaction of these different types of linguistic knowledge: syntax, semantics, discourse/pragmatics. In this chapter, we shall see that it is particularly in this latter domain that L1 children have been claimed to differ from L1 adults. This leads to several interesting predictions regarding the three-way comparison at the heart of the present study.

The L2 child and adult subjects in this thesis share the same L1, English. This language does not have word order alternations akin to scrambling, and rather than having two forms with associated differences in meaning, English has just one form which is ambiguous between specific and non-specific readings (cf. §2.5). In this chapter, it is argued that the acquisition of scrambling in Dutch represents a poverty-of-the-stimulus problem for the English-speaking L2er.

The chapter is organised as follows. It starts by briefly outlining the types of knowledge required for the acquisition of scrambling (§3.1). Subsequently, the bulk of the chapter reviews previous research on the production (§3.2) and the comprehension (§3.3) of scrambled objects in both L1 and L2 acquisition. Section 3.4 argues that if English-speaking L2ers are to acquire targetlike knowledge of scrambled objects, their interlanguage grammars must be constrained in the same way as L1 acquisition is. The chapter concludes with the predictions which will be tested in the remainder of the thesis.

3.1 Types of knowledge required for the acquisition of scrambling

The acquisition of scrambling involves the interaction of knowledge from several different sources: syntax, semantics and discourse/pragmatics. The required syntactic knowledge is the means to generate a direct object in the relevant position. For reasons outlined in Chapter 2, it is assumed that this involves movement. As detailed there, in certain contexts, the use of a scrambled form over a non-scrambled form can have consequences for interpretation. In other words, the acquisition of scrambling also involves the acquisition of semantic knowledge, namely that the scrambled object is of a different semantic type from the non-scrambled object. On Van Geenhoven's (1998) account, adopted by Krämer, the scrambled indefinite object is a free variable and the non-scrambled indefinite object predicative, whereas on de Hoop's (1992) approach, the difference is between a quantificational type and a predicate modifier (see §2.3.3). What is important on both of these approaches is that the semantic types in question are intricately related to each other and that the semantic type associated with the scrambled form is more complex. On Van Geenhoven's approach, this is because free variable indefinites need to be accommodated, and on de Hoop's analysis, the quantificational type is derived by type-shifting operations.

The type of discourse/pragmatic knowledge required for the acquisition of scrambling depends on the approach which is adopted, and this is presented in more detail in the relevant sections below. In the context of the present study, discourse and pragmatics are considered together, not because they should be viewed as one and the same thing, but because essentially, at least on two of the approaches under consideration here, this type of knowledge, which does not necessarily vary across languages (in the same way as syntax does, for example), forms a crucial part of the proposal in question, namely in the discussion of Schaeffer (§3.2.15) and Krämer (§3.3.1.1).[1] The details of which aspects of discourse and pragmatics are involved are presented in the appropriate places in this chapter. As the chapter progresses, we will see that such knowledge constitutes an important difference between children and adults and, consequently, a cross-group comparison of the kind carried out in this thesis will allow us to gain a better understanding of the interaction

[1] The terms 'discourse' and 'pragmatics' mean various things to various people. A detailed discussion of the difference between the two is beyond the remit of this thesis. Briefly, I consider pragmatics to refer to language use and discourse to refer broadly to inter-sentential relations. As noted in the main text, the crucial point is that these aspects of language as they are used in the work under review here are considered to operate in a similar fashion across languages.

between the discourse/pragmatic, semantic and syntactic components in the acquisition process.

The following two sections review previous research on respectively the production and the comprehension of scrambled objects in first L1 and then (adult) L2 acquisition.

3.2 The production of scrambled objects

Production studies address the question of when in a learner's linguistic development the scrambled form is used and whether this form is used in a targetlike fashion, that is, with the appropriate NP objects in the relevant contexts.

3.2.1 L1 acquisition

Several studies have investigated the production of scrambled objects in L1 Dutch and German, using both spontaneous and experimental data.[2, 3]

3.2.1.1 Barbier (2000)

Based on an elicited imitation task, Barbier (2000) argues that there is evidence for early competence in direct object scrambling (see also Penner, Tracy and Weissenborn 2000 on L1 German). She had 61 L1 Dutch children (2;8 – 6;3) each imitate eight sentences of approximately equal length containing definite NP objects scrambled over negation and over the sentential adverbs *nu* 'now' and *straks* 'later', illustrated in (1)-a and (1)-b, respectively.[4] The assumption behind such imitation tasks is that when a particular aspect of language is not part of a given subject's developing grammar, s/he will not be able to repeat it (as well) (Bley-Vroman and Chaudron 1994; Lust, Flynn and Foley 1996).

(1)	a.	Eva	mag	haar vinger	niet	in haar neus	steken
		Eva	may	her finger	not	in her nose	put
		'Eva may not put her finger in her nose.'					

[2] As noted in Chapter 2, German behaves the same as Dutch with respect to the properties under consideration here.

[3] Work on the acquisition of scrambling in Japanese and Korean is excluded from discussion here as this is usually concerned with scrambling to a pre-subject position. See, for example, Otsu (1994) and Lakshmanan and Ozeki (1996) on the acquisition of scrambling in L1 Japanese and Kang (2005) on L1 Korean.

[4] Barbier (2000) makes a distinction between scrambling across negation and scrambling across sentential adverbials, such as *vaak* 'often', which she classifies, respectively, as object shift and scrambling. This rather confusing use of terminology is based on Barbier's claim that unlike object shift, scrambling (over sentential adverbials) does not induce a change in meaning and hence the two processes are different. Although this holds for the types of sentences which she considers, that is, scrambling of definite NPs, there are, as the discussion in Chapter 2 illustrated, instances where scrambling over sentential adverbials *do* result in a change in interpretation, namely when the NP object in question is an indefinite.

b. Caroline mag de pop nu in het bad stoppen
Caroline may the doll now in the bath put
'Caroline may now put the doll in the bath.'

Barbier found that children of all ages imitated scrambled sentences in a targetlike fashion, with the older children more accurate than the younger children. Most of the errors involved either producing the object in a non-scrambled position (which she dubs a reconstruction error) or omitting the object altogether. On the whole, children were more successful at imitating sentences with scrambling over negation (as in (1)-a) than sentences with scrambling over adverbs (as in (1)-b), that is, they reproduced the object NP in the non-scrambled position for the sentences containing negation less frequently than for the sentences containing adverbs. This, Barbier claims, shows that children are sensitive to the interpretive constraints on scrambling over negation, that is, that scrambled definite NPs result in a sentential negation reading, whereas when definite NPs appear in a non-scrambled position a contrastive negation reading obtains. The children know that for the sentences containing adverbs, the position of the (definite NP) object is irrelevant for interpretation and hence, reproducing it in the non-scrambled position, which they frequently do, will not have an effect on meaning.

While Barbier's analysis is an appealing one, there are several weaknesses in her task which undermine her results and the claims she makes based on these results. Firstly, the nature of the task clearly puts the youngest subjects, the 2- and 3-year-olds, at a disadvantage: their more limited working memory capacities will undoubtedly affect how they perform on the task compared to the older children. In fact, across both conditions, the youngest group of subjects (2;8-3;3, n=7) scores on average just 0.08 out of a total possible two compared with 0.92/2 for the next age group (3;6-4;5, n=18). This rather poor performance also undermines Barbier's claim that there is evidence for early competence in scrambling. The claim that the younger children's performance was hindered by problems with the task is supported by the observation that a significant proportion of their responses involved something other than a correct answer, an omission or reconstruction (i.e. producing the non-scrambled order). This 'other' category accounted for approximately 54.5% of responses for scrambling across adverbs and approximately 70% of responses for scrambling across negation.[5] Second, the task included scrambled items only. The claim that the children's reconstruction responses to the sentences containing adverbs result from their knowledge of the interpretive constraints on scrambling would be more

[5] Barbier fails to provide any raw figures. These percentages are based on the data presented in her Tables 2, 3 and 4 and are calculated by adding up the average number of correct responses (Table 2), the average number of reconstruction responses (Table 3) and the average number of omissions (Table 4), subtracting this from the average total number of responses per condition (i.e. 2) and dividing this figure by the average total number of responses.

convincing if it could be shown that children equally alter the position of the object when imitating non-scrambled sentences, that is, if they were to scramble when they are asked to imitate non-scrambled sentences containing adverbs. Third, the sentences in this task were presented in isolation, that is, without any context. Any claims regarding the interpretation subjects attribute to a particular sentence should therefore be made with caution. Finally, there were no adult controls to establish targetlike response patterns.

3.2.1.2 Hoekstra and Jordens (1994), Eisenbeiss (1994)

We now turn to studies which deal with spontaneous production data. Generally, these studies have found that before they consistently produce scrambled objects in the appropriate contexts, L1 children pass through a "no scrambling" or optional scrambling stage. For example, Hoekstra and Jordens (1994) observed that at age 2;1, the Dutch-speaking child in their study did not scramble at all in obligatory contexts. Scrambling (over negation) was acquired gradually, with pronouns being the first objects to be scrambled, followed by proper names and definite NPs. According to the authors, at the age of 2;8, the only non-targetlike feature in the child's productions was her use of the negator *niet* with non-scrambled indefinite NP objects (e.g. *niet een pop* 'not a doll') instead of the suppletive form *geen* (e.g. *geen pop* 'no doll'; cf. §2.1.2). In her study of the spontaneous production of five German-speaking children, Eisenbeiss (1994) also observed that objects were initially produced in a non-scrambled position, before being scrambled optionally and subsequently, in a targetlike fashion. (2)-a provides an example of a targetlike scrambled pronominal object and (2)-b of a non-targetlike non-scrambled definite NP object. Both are taken from the same child, who had just turned two at the time of recording.

(2)	a.	simone	darf	des [= das (?)]	net [= nicht (?)]
		simone	is allowed	that	not
	b.	kann	nich	de knete	essen
		can	not	the dough	eat

(Eisenbeiss 1994:283, ex. 31 and 37)

Hoekstra and Jordens (1994) relate the acquisition of scrambling to the acquisition of the determiner system, noting that once determiners are produced, NP objects appear in their targetlike position with respect to scrambling. Eisenbeiss (1994) makes a similar but slightly more specific proposal, claiming that the acquisition of direct object scrambling is dependent on the syntactic activation of the AgrOP projection (the location of the scrambled object on her analysis), as manifested in the targetlike production of determiners, including targetlike accusative Case-marking.

3.2.1.3 Clahsen (1988)

Clahsen (1988) links the acquisition of scrambling to the acquisition of verb movement. He observes that in the early stages of acquisition, L1 German children initially treat sequences of a finite verb directly followed by negation, as in (3)-a, as syntactic islands (due to the obligatory affixation of Neg onto I°). As a consequence, no material, such as the object NP, *lala* 'dummy' in this example, is allowed to intervene between these two elements, that is, at this stage, scrambling is not possible in sentences with post-verbal negation.

(3) a. brauche nicht lala (Simone, 2;0.14)
need not dummy

b. Julia schere nich darf (Mathias, 2;9.07)
Julia scissors not may

(Clahsen 1988:130-131, ex. 115g and 117b)

This is not, however, because scrambling is not available at all, Clahsen claims. He observes that at the same stage, children *do* allow scrambling across negation when the verb is in final position, as in (3)-b. For Clahsen, the key to the acquisition of scrambling in sentences such as (3)-a is the acquisition of verb movement. Once children realise that finite verbs move from verb-final to verb-second position, they reanalyse the V_{+fin} plus negation sequences as derived structures, with the consequence that other elements are then allowed to intervene between the two, as in (2)-a (see also Clahsen 1990/1991; Weissenborn 1990).

Note, however, that all the examples which Clahsen provides to demonstrate scrambling with pre-verbal negation ((3)-b), compared with no scrambling with post-verbal negation ((3)-a), are taken from different children. There are no examples of both from one and the same child. Furthermore, the children who allow scrambling, that is, those who produce utterances such as (3)-b, are all older than the child who produces the utterances with no scrambling ((3)-a). Although such an age difference may be insignificant, it is telling that when the "no scrambling" child, Simone, produces targetlike scrambled utterances with post-verbal negation, as in (2)-a, she is around the same age as the children who produce the scrambled utterances in (3)-b with pre-verbal negation. This suggests that the utterances in (3)-a and (3)-b might not actually be taken from the same stage. In other words, early L1 German development might best be characterised by a "no scrambling" stage, too.

3.2.1.4 Interim summary

Although the spontaneous data used in the studies discussed above do, to a certain extent at least, allow the use of scrambled and non-scrambled objects to be evaluated for targetlikeness using the contextual information available in the transcripts, the relevant contexts, especially for scrambled (i.e. specific) indefinite objects, are very infrequent. In fact, none of the three studies explicitly discuss this latter type of object. Consequently, although the available evidence shows that L1 Dutch and German children start to use scrambled definite object NPs in their third year of life, the data regarding indefinite NP objects are less clear. Furthermore, given the limitations of spontaneous production data, we do not know whether, when L1 children are presented with a certain context, they will produce the targetlike scrambled or non-scrambled form. This was investigated, however, in an elicitation experiment by Schaeffer (2000). It is this experiment which will be (partially) replicated as part of the present study.

3.2.1.5 Schaeffer (2000)

Schaeffer's experimental study uses an elicited production task to test whether L1 Dutch children produce scrambled or non-scrambled forms when presented with certain contexts. The subjects were 49 monolingual L1 Dutch children, aged 2;4-6;10, and 23 adult native-speaker controls. In each of three conditions, subjects were presented with stories designed to elicit either a scrambled definite NP object, a scrambled (i.e. specific) indefinite NP object or a non-scrambled (i.e. non-specific) indefinite NP object. Example scenarios are presented in (4), (5) and (6), respectively.

(4) Schaeffer (2000): Example scenario for definite NP condition

Pluto:	Kijk,	een appel.	Bah,	die	lust	ik	niet	
	look	an apple	urgh	that	like	I	not	
	'Look, an apple! Urgh, I don't like that.'							
	Die	ga	ik	NIET	opeten			
	that	go	I	not	up-eat			
	'I'm not going to eat that.'							
Puppet:	De appel	gaat	Pluto	WEL	opeten			
	the apple	goes	Pluto		up-eat			
	'Pluto IS *going to eat the apple.'*							
Experimenter:	Nee he?	Wat	gebeurt	er	echt?			
	no right	what	happens		really			
	'No, right? What's really going to happen?'							
Targetlike response:	Pluto	gaat	de appel	NIET	opeten			
	Pluto	goes	the apple	not	up-eat			
	'Pluto is NOT *going to eat the apple.'*							
Non-targetlike response:	Pluto	gaat	NIET	de appel	opeten			
	Pluto	goes	not	the apple	up-eat			

(5) Schaeffer (2000): Example scenario for specific indefinite NP condition

Cookie Monster:	Kijk,	drie bananen:	1, 2, 3				
	look	three bananas	1, 2, 3				
	'Look, three bananas: 1, 2, 3'.						
	Twee	van	die bananen	ga	ik	NIET	opeten
	two	of	those bananas	go	I	not	up-eat
	'I'm NOT going to eat two of those bananas.'						
Puppet:	Twee bananen	gaat	Koekiemonster	WEL	opeten		
	two bananas	goes	Cookie Monster		up-eat		
	'Cookie Monster IS going to eat two bananas.'						
Experimenter:	Nee he?	Wat	gebeurt	er	echt?		
	no right	what	happens		really		
	'No, right? What's really going to happen?'						
Targetlike response:	Koekiemonster	gaat	twee bananen	NIET	opeten		
	Cookie Monster	goes	two bananas	not	up-eat		
	'Cookie Monster is NOT going to eat two bananas.'						
Non-targetlike response:	Koekiemonster	gaat	NIET	twee bananen	opeten		
	Cookie Monster	goes	not	two bananas	up-eat		

(6) Schaeffer (2000): Example scenario for non-specific indefinite NP condition

Ernie:	Mmmm,	ik	heb	zin	om	iets	te	eten
		I	have	desire		something	to	eat
	'Mmmm, I fancy eating something.'							
	Ik	kan	een snoepje	eten,				
	I	can	a sweet	eat				
	'I can eat a sweet …'							
	maar	dat	is	slecht	voor	je tanden		
	but	that	is	bad	for	your teeth		
	'…but that's bad for your teeth.'							
	Dus	dat	ga	ik	niet	doen		
	so	that	go	I	not	do		
	'So I'm not going to do that.'							

Puppet:	Ik	heb	het	niet	goed	verstaan	
	I	have	it	not	well	understood	
	'I didn't understand.'						
	Wat	gaat	Ernie	niet	doen?		
	What	goes	Ernie	not	do?		
	'What's Ernie not going to do?'						
Targetlike response:	Ernie	gaat	NIET	een snoepje	opeten		
	Ernie	goes	not	a sweet	up-eat		
	'Ernie is NOT going to eat a sweet.'						
Non-targetlike response:	Ernie	gaat	een snoepje	NIET	opeten		
	Ernie	goes	a sweet	not	up-eat		

Although scrambling over different kinds of adverbs was also tested, the younger children, who, as we shall see, are crucial to Schaeffer's account, failed to produce adverbs in these conditions, and hence, it was not possible to determine whether scrambling had taken place. The discussion here therefore concerns the data on scrambling over negation only, as does most of Schaeffer (2000). The results for both definite and indefinite NP objects are presented in Table 6.[6]

Table 6. Schaeffer (2000): Percentage of scrambled objects produced per age group and condition

Age group	n	Definite NPs	Indefinite NPs
2;4-2;11	7	30.4% (7/23)	33.3% (6/18)
3;0-3;11	13	72.2% (26/36)	56.3% (27/48)
4;0-4;11	11	81.6% (31/38)	57.1% (20/35)
5;2-5;11	10	76.5% (26/34)	58.5% (24/41)
6;0-6;10	8	82.8% (24/29)	56.7% (17/30)
Adults	23	96.3% (105/109)	66.3% (65/98)

Let us first consider the results for the definite NP condition. Adults almost always scrambled, whereas the 2-year-olds, as a group, only scrambled in approximately one third of items. By age 3, however, the children were significantly more adult-like. With the exception of the 2-year-olds, all the child groups did not significantly differ from the adults.[7] Turning to the indefinite

[6] Note that the definite NP data included definite NPs which were produced in other conditions as well as the condition designed to elicit them. Thus, when, for example, subjects produced a definite NP in the indefinite condition, this is included in the data in Table 6 in the first column.

[7] A comparable pattern was also observed for pronouns and proper names. The 2-year-olds scrambled personal pronouns at a rate of 33.3% (2/6), demonstrative pronouns at a rate of 16.7% (1/6) and proper names at a rate of 31.6% (6/19), compared with the 3-year-olds who

results, the first thing to note is that the figures given in Table 6 collapse responses made in both indefinite conditions (i.e. (5) and (6)), because even the adult controls failed to make a distinction between the two. Similar to the definite results, the indefinite results also show a leap between age two and three. As Schaeffer notes, however, given the problems with the indefinite scenarios, these results should be interpreted with caution. Closer examination of the specific indefinite scenario in (5) reveals that in addition to the scrambled response, in this scenario the non-scrambled response is in fact also possible. The non-scrambled response means something like 'the number of bananas which Cookie Monster is going to eat is not two' and this is true because he is only going to eat one banana. The fact that the adults sometimes failed to scramble in this condition therefore comes as no surprise. Putting the indefinites to one side, we now turn to Schaeffer's interpretation of the results, which is based on the definite NP data only.

Schaeffer concludes that in their acquisition of direct object scrambling, L1 Dutch children pass through two stages: first, they optionally scramble, producing both orders, and subsequently, they restrict their utterances (mainly) to the scrambled order only. She claims that children reach this second stage when, aged 3-4 years, they acquire a pragmatic principle which requires that hearer knowledge and speaker knowledge are always recognised as separate entities. The idea is that when children recognise hearer and speaker knowledge as separate entities, they can distinguish discourse-related NPs, that is, those which have been mentioned in the previous discourse, from non-discourse-related NPs, those which are part of the long-term shared knowledge between hearer and speaker. This distinction is crucial because, on Schaeffer's analysis, referential object NPs must be marked with either a [referential, discourse-related] feature or a [referential, non-discourse-related] feature in order to be able to scramble. If they are discourse-related, they scramble to [Spec, DiscP] and if they are non-discourse-related, they scramble to [Spec, RefP] (see §2.3.2.1 for more details and for a critical evaluation of this analysis of scrambling). In short, then, on this approach, if children do not separate hearer knowledge from speaker knowledge, they will not make the distinction between discourse-related and non-discourse-related and, consequently, they will not be able to scramble.

A vital component of Schaeffer's pragmatic principle is the word *always*: if children do not have the principle, they do not *always* recognise hearer and speaker knowledge as separate entities. This means that *sometimes*, they *will* recognise that these two things are not identical. When they do, they will be able to distinguish non-discourse-related from discourse-related NPs and consequently, they should mark the object with the relevant feature and

---

scored 95.2% (20/21) for personal pronouns, 77.8% (7/9) for demonstrative pronouns and 73.3% (22/30) for proper names. Note, however, that for some of these NP types, the numbers are very low.

scramble in the appropriate context. The claim that children do not *always* separate hearer knowledge from speaker knowledge is thus the source of optional scrambling. The existence of an optional scrambling stage is therefore a crucial consequence of Schaeffer's analysis. The data Schaeffer presents are group data only, however. In order to ascertain whether there really is an optional scrambling stage in early child L1 acquisition, it is necessary to examine individual results. The individual data for the 2-year-olds (kindly made available to me by J. Schaeffer) are as follows. One of the seven children does not produce any tokens. Of the remaining six, half scramble optionally (producing 1/3, 1/4 and 2/3 scrambled tokens, respectively) and the other half do not scramble at all (producing 0/3, 0/8 and 0/2 scrambled tokens, respectively). The observation that there are some children who appear to be in a "no scrambling" stage is consistent with Hoekstra and Jordens' (1994) and (my interpretation of) Clahsen's (1988) findings reported above. It is not, however, compatible with Schaeffer's analysis.

Central to Schaeffer's proposal is the claim that the separation of hearer knowledge from speaker knowledge has certain consequences in the child's developing grammar. That speakers alter their linguistic output according to the knowledge they do or do not attribute to their interlocutor is undisputed. That young children experience problems making this distinction and this is reflected in their language use is also relatively uncontroversial. What is not clear, however, is why separating hearer knowledge from speaker knowledge should lead to the distinction between discourse-related and non-discourse-related. Discourse-related and non-discourse-related NPs appear to be affected by the hearer/speaker knowledge distinction in a similar way. For example, in a household with a cat, *the cat* would generally count as a non-discourse-related NP, but if a member of this household were talking to someone who did not know that she had a cat, she may refer to it as *our cat* rather than *the cat,* that is, as a result of separating hearer knowledge from speaker knowledge, she would evaluate the long-term knowledge she shared with her interlocutor and alter her language accordingly.[8] Likewise for discourse-related NPs: when, in a conversation about films, for example, a speaker realises that her interlocutor is unlikely to know which film she is referring to, she will alter her language accordingly, perhaps saying *the film I saw last week in Utrecht*. The point is that while it is clear that making a distinction between hearer knowledge and speaker knowledge will have consequences for children's language, it is not clear that it should lead to the distinction which Schaeffer makes, namely between discourse-related and non-discourse-related; rather, the separation of hearer/speaker knowledge appears to straddle this distinction.

[8] It is also possible that the speaker will simply refer to *the cat,* assuming that the interlocutor will accommodate, that is, the interlocutor will infer that the speaker has a cat and that is the cat to which she is referring (cf. §2.3.3.2 for discussion of accommodation).

To summarise: On Schaeffer's account, the acquisition of the syntactic, semantic and discourse-related properties of scrambling crucially depends on the acquisition of a specific pragmatic principle. Once children have fully acquired this pragmatic principle, the syntactic consequences which Schaeffer claims this entails should follow automatically, that is, once children know that hearer knowledge and speaker knowledge are always separate, they should scramble when required. The principle which Schaeffer proposes is language-general: children acquiring any language will acquire it around the same age. The vital role played by a maturationally-determined, language-general principle on this account leads to interesting predictions for the L2 acquisition of scrambling, namely that any L2er older than the age at which this principle is acquired (3-4 years) should, all things being equal, be able to scramble. This prediction is spelled out in Chapter 5.

3.2.2 L2 acquisition

The acquisition of word order in L2 German and Dutch has received considerable attention over the years (e.g. Clahsen and Muysken 1986; 1989; duPlessis, Solin, Travis and White 1987; Jordens 1988; Schwartz and Tomaselli 1990). There are, however, hardly any studies which deal specifically with the acquisition of scrambling in these languages.[9]

3.2.2.1 Clahsen (1988)

As noted in the previous section, Clahsen (1988) found that L1 children initially fail to scramble (in certain contexts, if not in all). He also observes a "no scrambling" stage in the adult L2 acquisition of scrambling across negation in German by speakers of Italian, Spanish, Portuguese and Turkish. He reports that once these learners have progressed beyond an initial SVO stage with immediately pre-verbal negation, they pass through a stage where the finite verb and the non-finite verb are consistently separated from each other and negation occurs directly following the finite verb, to the left of the object. On Clahsen's analysis, the underlying grammar at this stage is still SVO and the non-finite verb moves to sentence-final position (Clahsen and Muysken 1986; 1989; cf. duPlessis et al. 1987; Schwartz and Tomaselli 1990). In a third phase, negation can be separated from the finite verb, that is, objects now appear in the scrambled position. Examples of each of these three stages are provided in (7) through (9), respectively.[10]

[9] For the production (and processing) of scrambled word orders in L2 Japanese, see e.g. Iwasaki (2003).

[10] The examples in (7) come from Turkish-speaking L2ers, whereas those in (8) are from Spanish and Italian speakers. It is not clear which L2ers produced the utterances in (9).

(7) Stage I. SVO with preverbal negation

a. aber nicht geben bei mir
but not give with me
'But they do not give it to me.'

b. aber nicht gefunden wohnung
but not found flat
'But I didn't find a flat.'

(Clahsen 1988:138, ex. 125d,e)

(8) Stage II. SVO plus verb displacement (SV_{+fin} $NegOV_{-fin}$)

a. ich will nicht mit sie gewohnt
I want not with them live
'I don't want to live with them.'

b. wenn ich glaube, ich kann nich(t) eine sache machen
if I believe I can not one thing do
'If I believe that I can't do a thing'

(Clahsen 1988:139, ex. 127a,e)

(9) Stage III. Separation of verb and negator/adverbs ($SV_{+fin}ONegV_{-fin}$)

a. ich versteh Michael wirklich nich(t)
I understand Michael really not
'I really don't understand Michael.'

b. ich wollte mich nich(t) bloßstellen
I wanted me not expose
'I didn't want to expose myself.'

(Clahsen 1988:139, ex. 129b,d)

An alternative analysis of these data, Clahsen notes, would be to propose the following stages of development, each of which is consistent with UG-governed principles:

(10) Stage I: Neg $_{VP}$[V…]
Stage II: Infl Neg $_{VP}$[…V] without scrambling
Stage III: Infl Neg $_{VP}$[…V] with scrambling

In stage I, the L2ers would have a head-initial VP with Negation adjoined to it. In stage II, the VP would change to head-final, but because scrambling is not yet available, negation would always occur directly after the finite verb in INFL, as in (8). Scrambling is finally acquired in stage III and, consequently, other constituents may appear between the finite verb and the negator, as in (9).

Clahsen rejects this analysis, however, giving the following three reasons. First, in stage II, the interlanguage grammar is supposed to have a head-final VP, but the L2ers' utterances at this point are still predominantly SVO. This, in itself, is not counter-evidence to the aforementioned analysis, however, because SVO is the target order in matrix sentences in German. To make this objection more convincing, it would have to be established that the L2ers' utterances are best characterised as $SV_{+fin}V_{-fin}O$. Clahsen's second objection also concerns stage II: according to the theoretical analysis which he adopts (namely Webelhuth and den Besten 1987), scrambling should automatically be available once the VP is set to OV. Hence, an OV stage without scrambling should be impossible. As we saw above, in the discussion Schaeffer (2000), this prediction does not seem to hold for L1 acquisition and hence it cannot be used as an argument against the analysis in L2 acquisition (see §3.3.2.1 for similar discussion regarding Neeleman and Weerman). Clahsen's third objection to the analysis given in (10) is that it cannot account for certain aspects of the data. Specifically, Clahsen observes that in stage II, utterances containing an NP object plus a PP occur exclusively with the scrambled order, (11)-b. The non-scrambled order, (11)-a, only occurs in stage III.

(11) a. V_i PP NP_{OBJ} $[e]_i$
b. V_i NP_{OBJ} PP $[e]_i$

Thus, for utterances containing an NP object and a PP, scrambling is available in stage II, whereas for the utterances with negation and a direct object, scrambling does not come in until stage III. According to Clahsen (1988:142) '[t]his contradiction cannot be resolved under the given grammatical analysis'. Without any quantitative evidence regarding the prevalence and exact distribution of these different orders, it is difficult to evaluate how real this contradiction potentially is. Furthermore, the only two examples of the non-scrambled order of stage III in (11)-a which Clahsen provides, replicated in (12), appear to be perfectly felicitous, because as far as is discernible, they contain non-specific indefinite NP objects.

(12) a. weil das ist für mich eine schlechte übersetzung
because that is for me a bad translation
'Because that is a bad translation for me.'

b. ich soll aus Italien eine apparat bringen
I should from Italy a device bring
'I am supposed to bring a device from Italy.'

(Clahsen 1988:140, ex. 129f,g)

This demonstrates that a more detailed examination of the distribution of the different kinds of NP objects is required before any firm conclusions can be drawn regarding the types of utterance given in (11). In short, if Clahsen's objections to the analysis given in (10) are to be considered valid, they require considerable elaboration (see also Jordens 1988:156-158 for relevant discussion).

Putting the problems with Clahsen's objections aside for a moment, let us briefly consider his alternative account for how scrambling is acquired. He claims that because the L2 data cannot be described in terms of possible UG-constrained grammars, an alternative, general learning strategy must be adopted instead. He proposes that L2 word order acquisition is constrained by the canonical word order strategy given in (13).

(13) In syntactic structures of the form

$_{XP}[X_1\ X_2 \ldots X_n]\ _{YP}[Y_1\ Y_2 \ldots Y_n]\ _{ZP}[Z_1\ Z_2 \ldots Z_n]$

in which each element $X_1\ X_2 \ldots X_n$ contains information on the internal structure of XP, no elements from YP, …ZP should be brought into XP

(Clahsen 1988:142, ex.132)

This strategy is based on the idea that when the canonical word order is violated, as a result, for example, of centre-embeddings, an utterance becomes more difficult to process. Clahsen claims that this limitation applies in adult L2 acquisition and that it is for this reason that in the earlier stages of development, word order patterns where the internal structure of a particular unit has been changed, due to scrambling, for example, do not occur. It is only once this 'processing constraint eases off' (Clahsen 1988:142) that word order patterns involving the movement of (part of) one constituent into the domain of another appear. These include scrambling and subject-verb inversion, which, in the data that Clahsen examines, emerge simultaneously, that is, in Stage III.[11] Clahsen claims that L2ers start to scramble once the general learning strategy/constraint 'eases off'. If the evidence for the 'easing off' of this strategy simply amounts to there being evidence for the use of word order patterns which it disallows, then the argument is rather circular. Independent evidence demonstrating that such a strategy exists, for example in other domains, and that it does indeed 'ease off' at a certain moment in time would strengthen Clahsen's claim.

To summarise: In one of the very few studies to specifically examine the L2 acquisition of scrambled objects, Clahsen (1988) shows that use of the targetlike scrambled form is preceded by a stage where, generally, the object remains in non-scrambled position. Without any detailed information regarding the type of object NPs occurring in each position and the contexts in which they appear, however, it is impossible to determine whether the L2ers'

---

[11] Presumably, several other properties of the target language should also be acquired at this point, but given that it is not obvious what the structure in (13) is actually supposed to represent, it is difficult to know what these other properties would be. More specifically, because the types of rules which Clahsen employs typically operate on the basis of linear order only, it is not clear whether XP, YP and ZP should be interpreted as maximal projections, in a hierarchical relationship, or as something else.

use of the scrambled form is targetlike in the sense that it is restricted to the relevant types of NP and contexts only. In order to find out whether this is the case, the present study will use an experimental method (similar to Schaeffer (2000)) to control for these variables.

Before turning to the specifics of this experiment (in Chapter 5), however, a review of the literature on the L1 and L2 comprehension of scrambled objects is in order.

3.3 The comprehension of scrambled objects

Comprehension studies have focused on the acquisition of scrambled indefinite objects. Specifically, child L1 studies have investigated whether the scrambled form is interpreted in a targetlike fashion, that is, whether scrambled indefinite object NPs receive a specific interpretation, whereas the only L2 comprehension studies which have been carried out to-date have predominantly dealt with whether the scrambled form is accepted by learners.

3.3.1 L1 acquisition

Several studies have investigated the acquisition of scrambled indefinite objects in L1 Dutch. For example, Bergsma-Klein (1996) observed that L1 Dutch children fail to assign a targetlike specific interpretation to scrambled indefinite objects. She used a truth value judgement task (see below) to determine whether children could correctly reject or accept scrambled and non-scrambled sentences such as (14) in the appropriate contexts.

(14)	a.	De jongen	heeft	een vis	twee keer	nagetekend
		the boy	has	a fish	two times	copied
		'The boy copied a (certain) fish twice.'				
	b.	De jongen	heeft	twee keer	een boom	nagetekend
		the boy	has	two times	a tree	copied
		'The boy copied a(ny) tree twice.'				

Bergsma-Klein found that on the scrambled sentences, none of the four age groups which she tested (four through seven years old) scored much higher than 50% correct. As Krämer (2000) notes, however, the number of variables in this study compared with the number of subjects means that its results must be treated with caution. For example, children were tested on four different types of sentences in two different conditions each and there were only seven subjects per group, with just two children in the 4-year-old group. In the following section, I review Krämer's study, which builds on the work of Bergsma-Klein, before briefly considering an alternative approach put forward by Philip (2003) and reviewing some of the numerous studies which have been carried out on the acquisition of scope in L1 English.

3.3.1.1 Krämer (2000)

As noted in Chapter 2 (§2.3.3.2), Krämer takes a semantic-pragmatic approach to scrambling. Following Van Geenhoven (1998), she assumes that scrambled indefinite NPs are free variables which are bound by the mechanism of accommodation. Krämer argues that the precise location at which this variable is bound is a function of bridging, and thus this latter process contributes to how precisely the scrambled indefinite is interpreted; whether, for example, it is considered to belong to a previously mentioned set (a 'source-set' guise, in Krämer's terms). Bridging (Clark 1977) is a more or less automatic process, driven by communicative requirements whereby anaphoric elements are related to antecedents available in the discourse context. Thus, in contexts where a source set is available, scrambled indefinite objects will automatically be linked to that set. Where no such set is available, the scrambled indefinite object will be interpreted specifically, that is, it will be assumed that the speaker had a particular referent in mind and this referent may or may not be known to the hearer.

On this approach, targetlike interpretation of scrambled indefinites requires knowledge of syntax, semantics and pragmatics. Each of these aspects is discussed in turn. Regarding syntax, the language-learning child (or adult) must to come to know that the scrambled position is the language-specific means of marking free variable indefinites. It was observed in the preceding section (§3.2) that although there is evidence to suggest that children initially fail to scramble direct objects, they start to produce definite and indefinite objects in a scrambled position at a relatively early age, around two or three years. From a relatively early age, then, L1 children acquiring Dutch and German demonstrate (some) knowledge of the syntax of scrambling. Krämer comments, however, that even though children may produce scrambled indefinites at this age, on the basis of production data alone it is impossible to know what interpretation they attach to them. Furthermore, as noted above, the specific indefinite condition in Schaeffer's production experiment is fundamentally flawed because in the specific indefinite condition, designed to elicit scrambled objects, a non-scrambled response was also legitimate (see also Su 2001 for similar discussion). This meant that even the adult native speakers did not make the relevant distinction. Note, however, that while Krämer is correct in pointing out the limitations of production data, if, in a similar (but appropriately modified) experiment, adult native speakers *did* make the relevant distinction in the appropriate conditions (that is, they scrambled in the specific condition and did not scramble in the non-specific condition) and children were observed to do the same, then in the absence of any convincing evidence to the contrary, one would be led to the conclusion that the interpretation which children attributed to these two forms was indeed targetlike.

Turning to the semantics of scrambling on Krämer's approach: as noted above, scrambled (free variable) indefinites are interpreted via the process of accommodation and non-scrambled (predicative) indefinites are incorporated into the verb (see §2.3.3.2 for details). To interpret predicative indefinites, the hearer only need pay attention to the information provided in the sentence in which the indefinite occurs. This contrasts with free variable indefinites, whose interpretation depends on the hearer being aware that a given utterance is part of a discourse, a context to which the present utterance adds new information. In order to establish whether children have the semantic prerequisites for the targetlike interpretation of these two types of indefinite, Krämer determines first whether children demonstrate knowledge of the predicative content of indefinites, as measured by the ability to use the same word to refer to different tokens, and second whether they are able to integrate discourse in language comprehension. Let us begin with predicative indefinites.

Krämer observes that the ability to use the same word to refer to different tokens is acquired very early, as evidenced by object-naming sequences such as the following:

(15)	Child:	Leeuw.						
		lion						
		'Lion'						
	Mother:	Ja.	Of	een poes				
		yes	or	a cat				
		'Yes. Or a cat.'						
	Child:	Ja						
		yes						
		'Yes.'						
	Mother:	Vind	't	meer	op	een poes	lijken	
		find	it	more	on	a cat	look	
		'It looks more like a cat to me.'						
	Child:	Dat	tijger					
		that	tiger					
		'That tiger.'						
	Mother:	Ja.	Een tijger					
		yes	a tiger					
		'Yes. A tiger.'						

(Krämer 2000:56, ex. 80)

For example, the utterance *dat tijger* in (15) predicates a property (that of being a tiger) of the object referred to with *dat*, thereby demonstrating the child's knowledge of predicative NPs. Children's use of intensional verbs and relational 'have', both of which require the semantic incorporation of their internal argument, are also used in a targetlike fashion from early on. An example of an intensional verb is given in (16).

(16)	Mother:	Wil	jij	wat	rozijntjes	hebben?
		want	you	some	raisins	have
		'Would you like to have some raisins?'				
	Child:	Ja				
		yes				
		'Yes.'				
	Mother:	Hè?				
		hm				
		'Hm?'				
	Child:	Wil	rozijntje			
		want	raisin			
		'Want raisin.'				
	(…)					
	Mother:	Jij	mag	rozijntjes	ja	
		you	may	raisins	yes	
		'Yeah, you can have some raisins.'				
	Child:	Dropje -	dropjes	hebben		
		liquorice	liquorice	have		
		'Liquorice, have liquorice.'				

(Krämer 2000:63-64, ex. 89)

As noted above, the semantics of free variable indefinites is more complicated because the process of accommodation, crucial to the interpretation of these NPs, requires discourse integration. It is this aspect of scrambling which Krämer claims causes children problems. For Krämer (2000:65), discourse integration involves two mutually inter-dependent sub-processes, namely (i) 'using the context in the structural interpretation of a particular linguistic element', and (ii) 'bringing utterances together into a whole that is in principle connected as to content'.[12] For accommodation to take place, it is enough for the context be part of the discourse representation. Recall, however, that on Krämer's analysis, the accommodation site of a free variable is identified via the process of bridging. For successful bridging to take place, the hearer needs to be aware of the types of information which may contribute to the context. This may include any preceding utterances, world knowledge and the visual/physical context.

Krämer notes that previous studies have observed the reflexes of children's poor discourse integration in language comprehension in various ways. For example, in a study on the comprehension of definite NPs in

[12] Discourse integration also requires a discourse representation. A discourse representation is 'a semantic representation of the discourse, which contains a representation of the individuals that play a role in the discourse [… and] the relations between the individuals that have been made explicit and whatever else is predicated of them' (Krämer 2000:65).

French, Karmiloff-Smith (1979) observed that children as old as eight failed to interpret the definite NP object, *the bell*, in (the French equivalent of) (17) anaphorically in around half the cases they were presented with (cf. just under 70% of non-targetlike responses for 3- to 5-year-olds and 17% and 10% for 9- and 10-year-olds, respectively).

(17) The boy touches a bell. And then the girl touches the bell.

In a series of similar pronoun resolution tasks, Yuill and Oakhill (1991) found that 7- and 8-year-old English-speaking children selected an implausible antecedent for a pronoun in 15% to 35% of cases. For example, when presented with the question in (18)-b, in the context of (18)-a, children provided the answer *Peter* rather than *John*.

(18) a. Peter lent his coat to John, because he was cold.
b. Who was cold, Peter or John?

Furthermore, Krämer notes that children committed more errors when memory load was increased by presenting the two clauses in (18) separately, that is, the first clause, *Peter lent his coat to John* , was removed (in the written presentation) when the second, *because he was cold*, was given. She concludes that these and other studies (e.g. Tyler 1983; Wykes 1981) demonstrate that children do have some discourse integration (because their responses are not completely non-targetlike) but that this is limited. This, coupled with Bergsma-Klein's results, leads Krämer to propose the Non-Integration Hypothesis:

(19) Krämer's Non-Integration Hypothesis:
The predicative interpretation of indefinites is acquired early.
The free variable interpretation is acquired later because it requires discourse integration.

In order to test this hypothesis, Krämer carried out two experiments on scrambling with L1 Dutch children.[13] The first experiment tested sentences such as:

(20) a.

De jongen	heeft	een vis	niet	gevangen
the boy	has	a fish	not	caught

'The boy didn't catch a (certain) fish.'

b.

De jongen	heeft	geen vis	gevangen
the boy	has	no fish	caught

'The boy didn't catch a(ny) fish.'

A standard truth value judgement paradigm was employed (Crain and Thornton 1998). Children were presented with a series of four pictures which told a story about a child and three objects. Two of these three objects were

[13] The complete description of these tasks is provided in Chapter 6, where they are replicated.

manipulated, leaving one remaining object in the final picture. Thus, in the case of the example given in (20), out of three possible fish in a pond, a boy catches two and leaves one behind. As a description of this context, the non-scrambled sentence in (20)-b is incorrect (because it is not true that the boy did not catch a fish) but its scrambled equivalent in (20)-a is correct (because there is a fish which the boy did not catch, namely the one left in the pond). There were six target items and six fillers (see §6.3.1.1 for details), with sentence type (scrambled vs. non-scrambled) and age as between-subjects factors. The results, along with the number of subjects in each age group who were tested on each of the two conditions, are presented in Table 7.

Table 7. Krämer (2000): Percentage of targetlike responses on "scrambling across negation" experiment

Age group	**n**	**Non-scrambled condition (target: rejection)**	**n**	**Scrambled condition (target: acceptance)**
4;0-5;6	6	100% (36/36)	7	16.7% (7/42)
5;6-6;10	6	100% (36/36)	8	22.9% (11/48)
6;10-7;10	–	–	11	10.6% (7/66)
Adults	4	100% (24/24)	6	100% (36/36)

All children correctly rejected the non-scrambled utterance as a description of the story given above, which is consistent with a specific interpretation of the indefinite object. Very few accepted the scrambled utterance in this context, however: just four out of the 26 children who were tested in the scrambled condition consistently accepted the scrambled utterance in all or all but one of the six trials. The observation that the non-scrambled, predicative indefinites caused the children no problems but the scrambled, free variable indefinites did is consistent with the Non-Integration Hypothesis (cf. (19)).

Krämer's second experiment involved an act-out task, which tested sentences such as:

(21) a. Je mag een potje twee keer omdraaien
you may a jar two times over-turn
'You may turn over a (certain) jar twice.'

b. Je mag twee keer een potje omdraaien
you may two times a jar over-turn
'You may turn over a(ny) jar twice.'

Here, subjects were provided with sets of four objects and asked to carry out a series of actions with them. For example, for the item illustrated in (21), subjects were presented with four jars. A targetlike interpretation would involve manipulating one jar only in response to (21)-a and one or two jars in response to (21)-b. (Recall that the set of circumstances in which the specific reading is true also verify the non-specific reading; in other words, in response to the non-scrambled utterance, it does not matter whether the object which is manipulated is the same or different.) There were six target items and five fillers and again, a between–subjects design was employed (see §6.5.1.1 for further details). The subjects in this task were different from those who completed the truth value judgement task. Their responses were categorised according to the number of objects which they manipulated (1-object response vs. 2-object response). The results are presented in Table 8.

Table 8. Krämer (2000): Percentage of targetlike responses on "scrambling across *twee keer* 'twice'" experiment

Age group	n	Non-scrambled condition		n	Scrambled condition	
		1-object response	2-object response		1-object response	2-object response
4;0-5;6	8	12.5% (6/48)	87.5% (42/48)	7	30% (12/40)	70% (28/40)
5;6-6;11.15	7	0% (0/42)	100% (42/42)	7	42.9% (18/42)	57.1% (24/42)
6;11.15-7;9	–	–	–	9	66.0% (35/53)	34.0% (18/53)
Adults	4	0%	100% (24/24)	6	91.6% (33/36)	8.4% (3/36)

In the non-scrambled condition, both children and adults generally chose to manipulate two objects (even though a 1-object response would not be incorrect).[14] In the scrambled condition, adults almost always manipulated one object only, whereas the children often provided a 2-object response in this condition, too. There is a developmental trend, with more correct responses with increasing age, but still, the oldest group of children remain significantly different from the adults. In the scrambled condition, all the adults had a 1-object response in all or all but one of the trials, whereas only nine out of the 23 children in this condition had a similar response pattern. Ten children consistently had a 2-object response pattern and the remaining four produced both responses.

The results from the truth value judgement task on scrambling across negation and the act-out task on scrambling across *twee keer* 'twice' are consistent with Krämer's Non-Integration Hypothesis: due to (hypothesised) limited discourse integration, L1 Dutch children fail to interpret scrambled indefinite objects in a targetlike fashion, even as old as age seven. Given that the oldest children tested by Krämer were still not targetlike (this is especially true for the negation experiment), we do not know at what age children consistently provide targetlike responses. This will be established with new

[14] Such a result (for the adults at least) is not entirely unexpected for the following reason: upon hearing the non-scrambled utterance in (21)-b, the adult native-speaker (the hearer) might (subconsciously) reason that were the 1-object response intended, the experimenter (the speaker) would have produced a scrambled utterance ((21)-a), because this would express such an intention unambiguously. The idea that the targetlike interpretation of scrambled objects involves reasoning about alternatives which the speaker could have produced is discussed in Chapter 7 (§7.4.2).

child L1 data in Chapter 6. Krämer suggests that children's errors result from their interpreting free variable indefinites as though they were predicates because these do not require accommodation.[15]

Note, however, that Krämer provides no independent evidence that those children who are non-targetlike have problems with discourse integration. Note, furthermore, that Krämer ultimately remains agnostic with respect to the underlying cause of children's limited discourse integration. She does, however, make two broad suggestions on this issue. These include (i) a lack of access to those components of the discourse representation necessary to interpret a given utterance and (ii) a default setting whereby children only pay attention to one utterance at a time. The latter, Krämer suggests, could be due to processing limitations and/or some kind of economy strategy or it could be innate. The former could also result from processing limitations, which prevent information from different sources (semantic, pragmatic, syntactic) from being combined, or from an incomplete discourse representation. The fact that Krämer leaves the cause of children's limited discourse integration so open is rather problematic when it comes to making specific predictions for L2 acquisition. It is clear, however, that processing limitations which stem from cognitive immaturity are different (at least in aetiology) from processing limitations caused by a low language proficiency level. In these terms, we therefore expect a difference between, on the one hand, adult L2ers, for whom any processing limitations can only be of the second type, and, on the other, child L2ers, whose processing limitations could be of either type, depending on the age of the subject at time of testing and proficiency level. It seems unlikely that L2 adults would resort to a default strategy where they pay attention to one utterance at a time only. Rather, it is expected that they would strive to integrate the discourse as they would in their L1. In the context of the negation experiment above, the simple vocabulary and visual support will surely facilitate this process.[16] Whereas such limitations would be more pervasive amongst the L2 children, it is only low-level L2

---

[15] Krämer dubs this type of error a 'semantic type' error. She also considers the possibility that the observed error patterns result from a 'pragmatic fit' error, that is, from a failure to construe the appropriate bridges between the free variable indefinite and its accommodation site. In the context of the two experiments reported on here, this could lead the child to identify a non-targetlike source-set, one which was either overly restrictive or not restrictive enough. Although this could explain the non-targetlike responses in the negation experiment (the child may have restricted the source set to one or both of the manipulated objects and evaluated the truth conditions for the test sentence with respect to these object(s) only), it could not explain the responses in the act-out task (because there, choosing an overly restrictive set of one object would result in a targetlike, 1-object response). Krämer does, however, find evidence for a 'pragmatic fit' error in some of her experiments on indefinite subject NPs.

[16] It could be argued that the simple vocabulary and visual support should also facilitate discourse integration in (L1/L2) children. This would depend on the cause of limited discourse integration.

adults who could plausibly be expected to have the processing limitations which may underlie problems with discourse integration.

To summarise: In a truth value judgement task on scrambling across negation and an act-out task on scrambling across the frequency adverbial *twee keer* 'twice', Krämer observes that children as old as seven regularly fail to access the specific interpretation for scrambled indefinite objects. Putting forward the 'Non-Integration Hypothesis', Krämer claims that children's problems with scrambled indefinites result from limited discourse integration. Discourse integration is crucial to a targetlike interpretation of the scrambled (free variable) indefinite because this allows the free variable's accommodation site to be identified. The fact that the ability to integrate discourse is not specific to a particular language leads to interesting predictions for child and adult L2 acquisition. The exact predictions regarding the child~adult differences in this domain are spelled out below (§3.5) and in Chapter 6.

3.3.1.2 Philip (2003)

Philip (2003) adopts a very different approach to the acquisition of scrambled indefinite NP objects. He argues that the specific interpretation of these NPs derives from a domain-general cognitive principle. This principle, stated in (22)-a, applies to a grammar when the linguistic condition in (22)-b is fulfilled.

(22) a. If there are two distinct cognitive objects X and Y of type A (e.g. linguistic forms) which both map to two distinct cognitive objects α and β of type B (e.g. meanings), then economise on information-processing resources by stipulating that X may only map to α and Y may only map to β.[17]

b. If a grammar G encodes the semantic difference between the specific and the non-specific reading of an indefinite NP of grammatical function Γ, then G must always optionally generate at least two distinct surface forms for an NP of grammatical function Γ.

(Philip 2003:6-7, ex. 5 and 6)

What this boils down to is that whenever a language encodes the specific/non-specific meaning distinction, its grammar will optionally generate two distinct forms and in order to maximise information-processing resources, a one-to-one mapping of form-meaning pairings will ensue. According to Philip (2003:4), languages such as Japanese and English do not grammaticise the specific/non-specific distinction and hence the cognitive principle in (22)-a does not apply.[18] Philip hypothesises that like UG, this principle, which is

[17] In the original, types A and B are represented with Japanese characters.

[18] Philip (2003:7) notes that Japanese allows scrambling but this is not associated with any interpretive effects; he suggests that this might result from a 'special instantiation of [(22)-a] that happens to occur in Dutch but happens not to occur in Japanese'.

optional in nature, is available from the onset of L1 acquisition.[19] Consequently, he claims, as many as half of the children acquiring Dutch should fail to apply it in the initial stages of acquisition, and those children who *do* apply it will soon stop doing so as a result of the non-targetlike positive evidence they will hear from their peers (that is, from those children who do not apply the principle). This, Philip argues, would lead to a situation in which no Dutch children have this one-to-one mapping rule and given that positive evidence produced by adults (or children) who do have the constraint could never lead to the acquisition of this rule, it can only be acquired via direct or indirect negative evidence. Acquisition via indirect negative evidence is a long and gradual process and hence, Philip predicts, on this 'Cognition-based Hypothesis', the acquisition of the targetlike interpretation of scrambled indefinites will be very late and very slow. This, Philip states, is the same developmental sequence which is expected for an adult Japanese L2er of Dutch; and such an L2er will 'have an advantage over an English learner of Dutch' (Philip 2003:10, fn. 11). This is presumably because English neither has scrambling (unlike Japanese), nor a grammaticalised specific/non-specific distinction. Whether English-speaking L2ers do take longer (or at least as long) as L1 Dutch children to acquire the target constraints will be discussed in Chapter 6 (§6.4.2.3). For the moment, we return to the prediction for the L1 child, namely that the acquisition of the targetlike interpretation of scrambled objects will be a slow and gradual process. (I discuss below whether the acquisition scenario on which this prediction is based is plausible.)

Philip tested this prediction on a total of 142 L1 Dutch children (aged 5;7-14;11), plus 40 adult controls.[20] Each subject was presented with an elaborate picture story about a boy, Kees, and his father, during which Kees kicked two different footballs (an orange one and a stripy one) into a goal.[21] Following a standard truth value judgement format (see above), a puppet then described the story using the following sentence:

(23)	Kees	heeft	een bal	dus	twee keer	in het doel	geschopt
	Kees	has	a ball	so	two times	in the goal	kicked
	'So Kees has kicked a ball in the goal twice.'						

The scrambled sentence in (23) is false on a target specific interpretation, because it means that there exists a ball which Kees kicked into the goal on two different occasions, but in the experimental scenario, Kees kicked two

[19] It is not clear to me whether the optionality arises from the optional application of the cognitive principle in (22)-a or from the nature of the linguistic condition in (22)-b.

[20] Amongst the child subjects, 27 started learning Dutch between birth and age four years. These children's (other) L1s included Moroccan Arabic, Turkish, English and French. These children were included in the final analysis because there were no significant main or interactive effects resulting from the variable 'bilingualism'.

[21] Several other experimental and control conditions were included but as these are not directly relevant to the current discussion, they are excluded from discussion here.

different balls into the goal. The results are presented in Table 9. Note that for reasons which will be discussed below, Philip only presents one trial of each sentence to each child and consequently the percentages in Table 9 represent the proportion of subjects who produced targetlike responses.

Table 9. Philip (2003): Proportion of subjects producing targetlike responses

	n	Scrambled condition (target = rejection)
6-year-olds	27	15%
7-year-olds	33	11%
9-year-olds	39	15%
12-year-olds	43	44%
Adults	20	87%

The youngest three child groups generally produce non-targetlike responses. The 12-year-olds are more targetlike, but they nevertheless remain significantly different from the adults (Kruskal-Wallis with H adjusted: $p \leq 0.0025$). Philip considers these results evidence of the slow and gradual developmental pattern one would expect if the acquisition of this particular property of Dutch proceeds on the basis of indirect negative evidence.

Philip claims that Krämer's Non-Integration Hypothesis does not really predict such a gradual acquisition pattern because if the targetlike interpretation of scrambled indefinite objects depends on targetlike discourse integration, a sudden dramatic improvement would be expected at the point at which this is acquired. Note, however, that this argument does not hold if discourse integration is acquired gradually. Philip notes this possibility (2003:16, fn. 15) but he considers it problematic because it is empirically indistinguishable from his 'Cognition-based Hypothesis'. Consequently, Philip's results cannot differentiate between these two proposals.

A further problem with Philip's prediction is the plausibility of the acquisition scenario on which it is based. Assuming for argument's sake, that the principle in (22)-a (or the rule in (22)-b – see fn. 19) is indeed optional, it seems reasonable enough to propose that if it were available at the onset of L1 acquisition, half of the children acquiring Dutch should apply it and half should not. What seems rather questionable, however, is that the non-targetlike utterances produced by the children that do not apply the rule should lead the other children, who originally did apply it, to stop doing so. If L1 children are so susceptible to the non-targetlike utterances produced by their peers, one might expect them never to reach adult levels of competence at all, or at least, they should be severely delayed for more properties of language in addition to the one discussed here. Furthermore, assuming that indirect negative evidence is necessary to acquire this particular property of Dutch, Philip fails to provide an explicit model of how exactly this should work. Language acquisition on the basis of indirect negative evidence entails that the language learner rules out as ungrammatical something she does not hear. It is not clear, however, how exactly this should work. As Pinker (1989:14 ff.) points out, it cannot be the case that language learners simply rule out any sentence which they have not yet heard because there are an infinite number of sentences for which this holds. Furthermore, it is not at all obvious how this would work anyway when it comes to <form, meaning> pairings. How does the language learner know which meaning, from a whole host of alternatives, she should rule out? Crucially, any language acquisition account based on indirect negative evidence needs to outline the exact circumstances in which a language learner should 'conclude that a nonwitnessed sentence is ungrammatical' (Pinker 1989:14) – or in the case of scrambling, that a nonwitnessed meaning is impossible – and this, as Pinker goes on to note, is more or less a restatement of the problem which the learner faces.

One final issue concerns a particular aspect of Philip's methodology. Philip (2003) questions the validity of Krämer's experimental findings, suggesting that her subjects' responses may have been biased as a result of the repeated measures design in her experiments. Specifically, referring to a phenomenon in the psycholinguistics literature referred to as a 'set perceptual response' or 'practice effect', he claims that when subjects are asked to respond to a particular type of stimulus on several different occasions (six times in Krämer's experiments), their responses will be affected by the fact that they have already heard such a stimulus once (or twice, etc.). Anything other than the first response can therefore not be considered as an independent measure of the subjects' knowledge. For this reason, Philip tests a large number of subjects with just one token in each condition.

There are two comments to make in this regard: First, it is not entirely clear what such a practice effect would look like: one could imagine that responses to previous items of the same type may have a strengthening effect, leading to consistent patterns of either acceptance or rejection, or it might be the opposite, causing the subjects to doubt their responses, resulting in a mixed pattern. It is not certain how one should deal with subjects who, for example, respond in a non-targetlike fashion to the first item of six but then consistently produce targetlike responses on the remaining items. Is this a practice effect or did these subjects simply make a mistake on the first item? On Philip's approach, we would not be able to see such potentially important patterns. Secondly, data based on one item only per subject cannot say anything about the stability of a given rule in a learner's grammar. This holds for L1 acquisition, but it is probably even more pertinent to L2 acquisition, where instability in interlanguage grammars is frequent. For this latter reason, a repeated measures design will be adopted here.

To summarise: Philip (2003) observes that the acquisition of the interpretive constraints on scrambled indefinite objects is a slow and gradual process, which is not yet completed at age 12. According to him, such delayed development cannot result from UG; rather, it follows from the (optional) application of a domain-general cognitive rule, which results in acquisition proceeding on the basis of indirect negative evidence only. Philip's approach was found to be lacking in two ways, however: (i) it fails to offer a specific model of indirect negative evidence demonstrating exactly how acquisition should take place, and (ii) the plausibility of the acquisition scenario which underlies it is questionable. Nevertheless, the empirical results will provide an interesting point of comparison for the data collected here.

3.3.1.3 Studies on the acquisition of scope in L1 English

The acquisition of scrambled indefinite objects involves the broader question of whether young children are able to access a specific or wide-scope interpretation for indefinite objects at all. Several researchers have investigated this question for the L1 acquisition of English as well as for other languages (Gualmini 2003; Lidz and Musolino 2002 (also on L1 Kannada); submitted; Musolino 1998; Musolino, Crain and Thornton 2000; Musolino and Lidz to appear; Su 2001 (also on L1 Chinese)). Given that these studies deal with English, which is the first language for the L2 subjects in the present study, and because they make claims which are also applicable to Dutch regarding the availability of specific readings for indefinites, a brief overview is provided (see also §6.6.1.3 and §7.3 for further discussion).

In a series of studies, Musolino and Lidz (Lidz and Musolino 2002; Musolino 1998; Musolino et al. 2000; Musolino and Lidz to appear) have observed that young children (around four years old) acquiring English consistently interpret NPs in their surface scope position. Object NPs such as those given in (24) are interpreted as taking narrow-scope with respect to negation and the subject NPs in (25) receive a wide-scope interpretation, whereas in the adult grammar, both these types of sentence are ambiguous. Thus, in a situation where Cookie Monster ate two cookies and left two cookies, children typically reject (24)-b, that is, they prefer to interpret the object in its surface position, within the scope of negation. In contrast, in a context where there are two butterflies who go to the city and two who do not, children will accept sentences such as (25)-b, that is, they prefer to interpret the subject in its surface position, which in this case is outside of the scope of negation. Musolino *et al.* (2000) dub this pattern the 'Observation of Isomorphism'.

(24) a. The smurf didn't buy every orange
b. Cookie Monster didn't eat two cookies
c. Donald didn't find some guys

(25) a. Every horse didn't jump over the fence
b. Two butterflies didn't go into the city

Lidz and Musolino (2002) claim that children's preference for an isomorphic interpretation follows from their use of the hierarchical relations of c-command: as the representation in (26) shows, the indefinite object NP *two cookies* in (24)-b, is c-commanded and hence in the syntactic scope of negation.[22]

(26) [$_{IP}$ Cookie Monster [$_{I}$ didn't [$_{VP}$ [$_{V}$ eat [$_{NP}$ two books]]]]]

As Krämer (2000) notes, such an analysis is at direct odds with the L1 Dutch findings. An isomorphic interpretation of a sentence with an indefinite object scrambling across negation would lead to a targetlike, wide-scope (i.e. specific) interpretation, but this is not what Dutch children actually do: they consistently assign scrambled indefinite objects a non-isomorphic, narrow-scope (i.e. non-specific) interpretation.

Similarly to the aforementioned studies, Su (2001) also found that L1 English children (n=18 ; 3;10-6;2, mean 5;0) consistently assigned indefinite NP objects embedded under negation a non-specific interpretation in contexts where a specific interpretation was possible. In particular, subjects in a truth value judgement task were presented with the sentence given in (27), with a scenario where Mickey rode two out of three available dogs. (This is similar to Krämer's fish story discussed above in §3.3.1.1.)

(27) Mickey didn't ride a dog

In such a context, this sentence is ambiguous for the adult native speaker: on a specific reading it is true because there is one dog which Mickey did not ride, and on the non-specific reading, it is false because Mickey *did* ride a dog. This is reflected in the 50-50 split observed in the adults' (n=30) responses (60/120), that is, they accept/reject the specific reading in half of the trials. The individual data (kindly made available to me by Y. Su) show that there is considerable variation amongst subjects, however. Thirteen of the 30 native adults (43.3%) accept the specific reading in all or all but one of the four trials,

---

22 Similar observations were made for children acquiring the Dravidian language, Kannada. Unlike English, in this language, c-command relations do not coincide with linear order, and hence Lidz and Musolino (2002; submitted) use data from L1 Kannada children to tease apart these two properties. They find that children interpret numerically quantified NP subjects and objects on the basis of c-command relations rather than linear order.

five (16.7%) accept it in half of the trials and the remaining 12 subjects (40%) rejected it in all or all but one of the trials.

As a group, the children only accepted sentences such as (27) in 33.3% (24/72) of trials (cf. 50% for the adults). In other words, as a group, they assigned the indefinite a narrow-scope reading more frequently than the adults. They motivated their responses by stating, for example, that 'Mickey did ride two dogs'. As with the adults, the response patterns varied across individual subjects: a third of the 18 subjects accepted the specific reading in all or all but one of the trials, one subject accepted it in half of the trials and the remaining 11 subjects (61.1%) rejected the specific reading in all or all but one of the trials. The children thus differ from the adults in that there are more individual subjects who exhibit a pattern of rejection. Proportionally, however, the difference between the children and the adults is rather small: 40% of adults (12/30) consistently rejected the specific reading compared with 61.1% of the children (11/18). This suggests that most of these children must therefore be targetlike.[23]

Miller and Schmitt (2004) replicated Su's experiment and found that as a group, their 29 adult native-speaker participants accepted sentences like (27) in 65.5% (76/116) of trials in a context where both specific and non-specific readings where available (as above). When this percentage was broken down across the four trial sentences, however, it was observed that there was considerable inter-trial variation. The mean percentages for the four different test sentences ranged from 34.5% (10/29) for 'Troy didn't buy a hat' to 86.2% (25/29) for 'Denny didn't eat a cookie'.

The results from this and Su's study are particularly interesting in the context of the present study because they provide information about what English-speaking L2ers of Dutch might transfer from their L1 when confronted with indefinite objects in negative sentences in Dutch. These results also provide empirical support for the observation made in §2.2 that sentences such as (27) are in general ambiguous for adult native speakers, although as the individual data show, this ambiguity varies across sentence types/scenarios and across individual speakers. This means that when the English-speaking L2ers of Dutch in the present study are tested using a comparable experiment, if they accept the specific reading of the scrambled indefinite object, it cannot automatically be assumed that they have targetlike knowledge of the relevant interpretive constraints, as such a response could potentially be the result of L1 transfer. What we should find, however, is that when subjects accept this reading for the right reason, the proportion of subjects with targetlike responses should increase with proficiency. For the L2

[23] It is possible, however, that children and adults may reject the specific reading for different reasons: the children may do so because they cannot access the specific reading at all, whereas the adults may (for some reason or another) not be able to access it in the particular circumstances of the experiment or they simply prefer the non-specific reading.

children, the situation is slightly more complicated, however. Su's results suggest that in the current study some of the younger L2 children (when age is measured at time of testing) might not be targetlike in their L1, English, which could mean that they might not be able to access the specific reading at all. On such a scenario, failure to accept the specific reading on the Dutch task might be a consequence of a general inability to access this reading, rather than the result of L1 transfer or a lack of targetlike L2 knowledge.

To summarise: Several studies on the acquisition of scope in L1 English have shown that 4-year-old children often fail to accept a specific interpretation of indefinite object NPs embedded under negation. It has been claimed that this is because children initially interpret object NPs in their surface scope position. Such a claim is, however, at odds with the L1 Dutch data. The L1 scope studies also highlight that sentences containing indefinite object NPs under negation are ambiguous for native speakers of English. This has important implications for the role of L1 transfer in the context of the present study.

3.3.2 L2 acquisition

The (very) few studies dealing with the L2 comprehension of scrambled objects have only investigated whether adult L2ers accept scrambled orders. Despite the increasing attention devoted to the syntax-semantics interface in L2 acquisition research, no study has investigated the particular property of Dutch under consideration here.

3.3.2.1 Neeleman and Weerman (1997)

Neeleman and Weerman (1997) investigate the acquisition of scrambling as part of their OV/VO parameter (cf. §2.3.2.2). They used an untimed grammaticality judgement task to test whether English-speaking L2ers of Dutch knew the basic OV word order of Dutch and the properties of this language which, on their proposal, fall out of the OV setting of the aforementioned parameter, that is, scrambling, Exceptional Case Marking, the distribution of verb particles and the extraction from objects of complex predicates.

The subjects were 14 English-speaking 13-14-year-olds who had been learning Dutch for between one and eight years in an international school in The Netherlands where English is the language of instruction.[24] They received three hours of Dutch language instruction per week (plus homework); no information is provided regarding the amount of contact they had with Dutch outside the classroom. Subjects were presented with written stimulus sentences and asked to judge whether they were 'correct' or 'incorrect', and if they were

[24] On the criteria adopted here, at least one of these subjects would be classified as a child L2er. (Neeleman and Weerman do not provide any information about individual subjects.)

incorrect to correct them. Examples of the scrambling stimulus sentences are provided in (28). Note that both sentence types are grammatical.

(28) Non-scrambled stimulus sentence

a.	Natascha	moet	iedere	dag	haar	huiswerk	maken
	Natascha	must	every	day	her	homework	do

'Natascha has to do her homework every day.'

Scrambled stimulus sentence

b.	Olga	heeft	het	boek	langzaam	gelezen
	Olga	has	the	book	slowly	read

'Olga read the book slowly.'

(Neeleman and Weerman 1997:168)

The authors adopt an acquisition criterion of 75% targetlike responses. For the scrambling sentences, 11 out of the 14 subjects fulfilled this criterion. There was no clear relationship between this and the other sentence types, including OV word order. The observation that the various properties of the OV/VO parameter do not cluster together in the L2 data lead Neeleman and Weerman to argue that (adult) L2 acquisition cannot be constrained by UG. However, Neeleman and Weerman also fail to observe a clustering effect in the L1 spontaneous production data they examine. According to these authors, though, this is a result of independent factors, such as computational complexity. They propose that crucially, the error patterns observed in the L1 spontaneous production data are consistent with the idea that the acquisition of the various properties associated with this parameter presents a single learning task. Logically, however, such reasoning could also be applied to the L2 data, too. This disparity between the L2 and the L1 data and how they are analysed is just one of the many problematic aspects of this study.

The data Neeleman and Weerman examine are of completely different types in the two cases and the criteria they use to determine acquisition are also different for the two groups of learners. The L1 data are spontaneous production data and emergence counts as acquisition, whereas the L2 data are judgement data and 75% correct is the acquisition criterion (see White 2003 for similar objections). Their claim that scrambling is optional from the earliest stages of L1 acquisition is partly based on Schaeffer's experimental data. It was shown above, however, that this claim only holds for half of the individual children in the youngest age group. The other half consistently fail to scramble, which goes against Neeleman and Weerman's predictions.

A further problem with Neeleman and Weerman's study is that the design of the L2 task is rather limited and it may also have biased the results. In the grammaticality judgement task, subjects were asked to judge contextless scrambled and non-scrambled sentences. Although this is potentially informative with respect to whether a certain construction is allowed in the interlanguage grammar (see also Hopp 2005 discussed below), in order to obtain a complete picture of the acquisition of scrambling, it is arguably necessary to assess whether L2ers also know the interpretive constraints on scrambled objects. Granted, the task used (definite) object NPs which, in the target grammar, can appear in either position without any resulting change in meaning and hence it was not designed to test interpretation. This does not mean, however, that a developing interlanguage grammar will necessarily also allow both options to mean the same thing. It is possible that subjects conjured up their own contexts or applied certain intonation patterns such that the given sentence was rendered (in)felicitous, and hence they judged the sentence to be (un)acceptable, not because of its form but because of the meaning they attributed to it. Given that the task was written and contextless, such a possibility cannot be ruled out. The task was also untimed and it asked the (instructed) subjects to judge sentences as 'correct' or 'incorrect'. Both of these aspects could encourage the use of metalinguistic knowledge; the results from such a task therefore might be relatively uninformative with respect to the status of such sentences in the L2er's interlanguage grammar. Finally, as the authors themselves note, the fact that both the scrambled and non-scrambled sentences were grammatical means that the subjects' relatively successful performance (especially compared to other sentence types) might result from a yes-bias, rather than from grammatical knowledge.

The fundamental flaws in Neeleman and Weerman's study undermine their results. To improve upon their task, it would be necessary to (i) use the same methodology with all subject groups being examined, (ii) provide a context and intonation for stimulus sentences and (iii) include stimulus sentences which require a 'no-response'. Each of these requirements is fulfilled in the tasks employed in the present study.

3.3.2.2 Hopp (2002; 2005)

The acceptability of scrambled sentences is also tested in a recent study by Hopp (2002; 2005). Building on work by Schreiber and Sprouse (1998), Hopp investigates the L2 acquisition of scrambling in German by speakers of English and Japanese with advanced and near-native German proficiency. His study is concerned with scrambling to a pre-subject position and the graded acceptability of various optional movement operations involving scrambling and topicalisation and remnant movement. It is therefore not directly relevant to the present study and as such will not be discussed any further, except to note that Hopp shows that the advanced non-native subjects in his study accept (licit) scrambled orders and their judgements reflect the gradedness of acceptability shown by natives. For the English-speaking L2ers, the acquisition of such knowledge constitutes overcoming a poverty-of-the-stimulus problem. It is to the issue of the L2 poverty-of-the-stimulus and the syntax-semantics interface to which we now turn.

3.3.2.3 Studies on the syntax-semantics interface and the L2 poverty-of-the-stimulus

Pioneering research on the syntax-semantics interface in L2 acquisition has been carried out by Laurent Dekydtspotter, Rex Sprouse and colleagues (Dekydtspotter, Sprouse and Anderson 1997; 1998; Dekydtspotter, Sprouse, Swanson and Thyre 1999; Dekydtspotter, Sprouse and Thyre 1999/2000; Hopp 2005; see also Juffs 1996; Marsden 2004; Montrul and Slabakova 2003; Slabakova 2003; White 1998). Focussing predominantly on English-French interlanguage in adult L2ers, they have investigated the acquisition of a variety of phenomena including the process/result distinction, relative clauses, quantifier/tense interactions and *combien* 'how many' extractions. The latter will be used to illustrate their approach in the first part of this section. In the second part, I turn to recent work on the syntax-semantics interface in L2 Japanese by Marsden (2004).

3.3.2.3.1 Dekydtspotter, Sprouse and Thyre (1999/2000)

French allows the interrogative numeral determiner *combien* 'how many' to appear either together with or separated from its nominal complement. These so-called continuous and discontinuous constituents differ in interpretation. This is illustrated in (29).

(29) a.

Combien	de livres	est-ce	que	tous	les étudiants	lisent?
how many	of books	is-it	that	all	the students	read

'How many books do all the students read?'

b.

Combien	est-ce	que	tous	les étudiants	lisent	de livres?
how many	is-it	that	all	the students	read	of books

How many books do all the students read?'

In a situation in which one student, Graham, reads *The Tax Inspector*, *Oscar and Lucinda* and *Jack Maggs* and another student, Lorna, reads *Oscar and Lucinda*, *Jack Maggs* and *Illywhacker*, the answer to the question in (29)-a is either two on the wide-scope or collective reading (there are two books which both Graham and Lorna read, namely *Oscar and Lucinda* and *Jack Maggs*) or three on the narrow-scope or distributive reading (the number of books per student). The question in (29)-b, on the other hand, allows the narrow-scope/distributive reading only. According to Dekydtspotter *et al.* (1999/2000), the unavailability of the wide-scope reading in (29)-b results from the interaction of various syntactic and interface constraints (Diesing 1992; Obenauer 1984/1985), the details of which need not concern us here.

For the English-speaking L2er, the acquisition of this particular property of French represents a poverty-of-the-stimulus problem. There is nothing in the TL input or the L1 to prevent the L2er from assuming that the (much less frequent) discontinuous pattern is simply a rewrite of the continuous pattern because (i) the L1 equivalent, the continuous form given in the translation gloss in (29), allows both narrow and wide scope readings and (ii) the situations in which the discontinuous *combien* interrogative is true are a proper subset of the situations in which the continuous *combien* interrogative is true. Thus, Dekydtspotter *et al.* argue, if L2ers demonstrate knowledge of this property of French, which furthermore is not covered in the L2 French classroom, this must result from the L2 hypothesis space being severely constrained, in a similar fashion to L1 acquisition.

Beginner/intermediate (n=21) and advanced (n=19) English-speaking adult L2ers of French were tested using a truth value judgement task. Subjects were presented with continuous and discontinuous *combien* questions in two different contexts, one consistent with a narrow-scope answer to the question and another consistent with a wide-scope answer. The beginner/intermediate group failed to make a distinction between the continuous and discontinuous questions in the wide-scope condition (accepting them in 42.2% of items in both conditions; $p > 0.5$), whereas like the native controls, the advanced L2ers did make such a distinction, consistently rejecting the wide-scope reading for the discontinuous question more frequently than for the continuous question (75.0% vs. 52.4%; $p < .0005$). Such a distinction, Dekydtspotter *et al.* claim, 'could not feasibly be acquired without a restricted relation between levels of syntactic and conceptual structure representations' (Dekydtspotter et al. 1999:170).

3.3.2.3.2 Marsden (2004)

In her recent PhD dissertation, Marsden (2004) investigates a similar poverty-of-the-stimulus problem. She examined the availability of object-wide scope in Japanese sentences with an existentially-quantified subject and universally-quantified object. In English and Chinese, such sentences, illustrated for English in (30), allow both a subject-wide (some > every) reading and an object-wide (every > some). The subject-wide reading can be paraphrased as 'there is some person *x* such that *x* read every book' and the object-wide reading as 'for every book *y*, there is some person who read *y*'. In Japanese and Korean, however, the only possible reading is the one which corresponds to linear scope (some > every), as shown for Japanese in (31).

(30) Someone read every book (some > every, every > some)

(31) Dareka-ga daremo-o semeta (some > every, *every > some)
someone-NOM everyone-ACC criticised
'Someone criticised everyone'

Marsden employed a comparative design, investigating the acquisition of Japanese in three different L1 groups, English, Chinese and Korean. As noted above, English and Chinese pattern differently from Japanese and Korean. For the L1 English and L1 Chinese speakers, the acquisition of the scope rigidity exhibited in Japanese presents a poverty-of-the-stimulus problem: there is nothing in the target language input to dissuade these L2ers that the object-wide reading, which they are assumed to transfer from their L1, is not possible in Japanese. Furthermore, this property of Japanese is not covered in the L2 classroom (Marsden 2004:12, fn. 15).[25] For the L1 Korean speakers, there is no such poverty-of-the-stimulus problem, however, as their L1 exhibits the same pattern. By comparing these different groups of L2ers, Marsden's innovative study not only informs the poverty-of-the-stimulus debate in L2 acquisition, but it also documents the role of L1 transfer in the acquisition of a target language property at the syntax-semantics interface. Given that the behaviour of the L1 English speakers is most relevant to the present study, the remaining discussion is restricted to the results for this group only.

A picture-sentence matching task was employed to determine which readings the L2ers (and native adult controls) allow for sentences such as (31). For example, in order to determine whether a subject-wide scope reading (some > every) was allowed, subjects were asked whether the sentence in (32) matched a picture where one girl is stroking three cats. For the object-wide scope reading (every > some), they were presented with a picture where three different children were stroking one cat each. To demonstrate knowledge of the target interpretive constraints, L2ers should accept the former picture-sentence combination and reject the latter one.

(32)	Dareka-ga	dono neko-mo	nadeta	(some > every, *every > some)
	someone-NOM	every cat-QPt	stroked	
	'Someone stroked every cat.'			

The results were as follows. The intermediate level English/Japanese L2ers (n=20) allowed both readings, whereas the advanced English/Japanese L2ers (n=9) were generally targetlike, that is, they correctly rejected the object-wide scope reading. These two results demonstrate (i) that L2ers whose L1 allows scope ambiguity in sentences containing an existentially-quantified subject and a universally-quantified object transfer this to their interlanguage grammar, and (ii) that such L2ers are eventually able to overcome the poverty-of-the-stimulus which the acquisition of such sentence types entails.

Marsden also examined the L2ers' knowledge of scrambling. In Japanese, when a universally-quantified object is scrambled across an existentially-quantified subject, as in (33), both the subject-wide *and* the object-wide readings are available. In other words, scrambling has the effect of 'freeing up' the scope rigidity observed in non-scrambled sentences, such as (31).

---

[25] Marsden's subjects were all classroom learners, although some had been exposed to Japanese in Japan.

(33) daremo-o Dareka-ga t_i semeta (some > every, every > some)
everyone-ACC someone-NOM criticised
'Someone criticised everyone'

The results show that (on the whole) the L2ers from all three language backgrounds pattern as the natives: they accept both subject-wide and object-wide readings on the scrambled sentences and the acceptance rate of the object-wide reading for the scrambled sentences is always (and mostly statistically) higher than for the non-scrambled sentences.

3.3.2.3.3 Summary

Taken together, the Marsden and the Dekydtspotter *et al.* studies demonstrate that adult L2ers are able to overcome poverty-of-the-stimulus problems instantiated at the syntax-semantics interface. More specifically, these studies show that in the L2 acquisition of a restricted scope interpretation associated with a particular form, L2ers whose L1 has only one, scopally ambiguous form only, initially transfer this ambiguity to their interlanguage grammar. Ultimately, however, they are able to acquire the scopal restrictions instantiated in the TL. The present study seeks to build on and contribute to the increasing body of L2 research on the syntax-semantics interface in two ways: first, by examining a comparable acquisition scenario in a different L1/TL pair, and second, by investigating the syntax-semantics interface in child L2 acquisition as well as in adult L2 acquisition. In the following section, I outline how the L2 acquisition of scrambling in Dutch by native speakers of English presents a poverty-of-the-stimulus problem.

3.4 The L2 acquisition of scrambling as a poverty-of-the-stimulus problem

This section outlines how the L2 acquisition of scrambling by native speakers of English constitutes a comparable acquisition scenario to the studies reviewed in the preceding section. In those studies, the L2ers were faced with the following learnability problem: they had to move from an L1 grammar where one (syntactic) form (or string) allows both wide- and narrow-scope readings of the object NP to a target language grammar where there are two forms and one of these disallows the wide-scope reading for the object. This was the canonical (quantified) subject – (quantified) object – verb order in Marsden's study on L2 Japanese and the discontinuous form of the *combien* question in Dekydtspotter *et al.*'s study on L2 French.

The L2ers in the present study are faced with a comparable – though not identical – acquisitional task: the L1 (English) has one (syntactic) form, namely the order given in (34), which allows two different interpretations for the object, whereas the TL has two different forms, the scrambled and non-scrambled, and one of these, namely the scrambled order, is restricted in its interpretation.[26] This is illustrated for scrambling across *twee keer* 'twice' in (35) (see §2.5 for more details).

(34) The girl threw a ball twice / two times

(35) a. Het meisje heeft een bal twee keer gegooid
the girl has a ball two times thrown
'The girl threw a (certain) ball twice.'

b. Het meisje heeft twee keer een bal gegooid
the girl has two times a ball thrown
'The girl threw a(ny) ball twice.'

The L2ers have several different sources of knowledge at their disposal, including the L1, instruction and the target language input. It is not possible to acquire the contrast illustrated in (35) from these sources alone, however. The L1, English, cannot provide the necessary information because as noted in §2.2 (cf. also §2.5.1), English does not have scrambling and equivalent sentences are ambiguous between a specific and a non-specific interpretation.[27] The relevant

---

[26] (British) English also has the marginally acceptable form given in (i). This is interpreted as (34).

(i) The girl ?twice / *two times threw a ball

[27] There are, however, circumstances in which an indefinite object NP (generally) takes wide-scope with respect to a scope-taking operator, namely in the double-object construction. Here, the scopal relations are read off the linear order of the indefinite indirect object and quantified direct object. Thus, the sentence in (i)-a is consistent both with a situation where Jean gave each toy to a different child and with a situation where she gave all the toys to one and the same child. The sentence in (i)-b, however, is only consistent with the latter situation. In other words, the only reading possible is one where the indefinite indirect object takes scope over the quantified direct object.

(i) a. Jean gave every toy to a child
b. Jean gave a child every toy

It has been suggested (W. O'Grady p.c., 7 December 2004) that this contrast may facilitate the English-speaker's acquisition of the relevant contrast in Dutch. There are two reasons why this seems unlikely, however. First, on most analyses of the double-object construction, the direct and indirect objects in (i)-b either both move (for QR purposes, Bruening 2001; or to check their features in the specifier position of a higher functional projection, see e.g. Radford 1997) or both are in their base position (e.g. Pesetsky 1995). If there is movement (at any level), therefore, it is not this movement per se which can be linked to the lack of a narrow-scope reading for this object relative to the quantified direct object. It is therefore difficult to see how this would transfer to scrambling in Dutch. If there is no movement in English (i), scope relations are presumably read off the linear order (or c-command relations) on the non-specific reading and using QR on the specific reading. That English-speaking L2ers might interpret scrambled objects in Dutch on the basis of surface scope is of course a logical possibility information is also not the focus of instruction in the L2 classroom. This was confirmed by Dutch language teachers as well as consultation of grammar books. Instruction is often given on (general) word order and the use of *geen* 'no/not any', but the interpretive facts or constraints under investigation here are not provided. In fact, most of the (native speaker) teachers consulted were not consciously aware of the relevant distinction.[28]

Concerning scrambling across the frequency adverbial *twee keer* 'twice', the target language input will provide evidence that the non-scrambled form, such as (35)-b, is consistent with situations where two different objects or one and the same object is manipulated (the non-specific interpretation) and that the scrambled form is consistent with situations where one object is manipulated (the specific interpretation). It will provide no direct, positive evidence that the scrambled form is inconsistent with a situation in which two different objects are manipulated, however. This means that, as in Marsden's study, in order to acquire the target language interpretive constraints, the English-speaking L2ers of Dutch would have to notice that a particular form~meaning pairing does not occur and from this non-occurrence induce its ungrammaticality.[29] In order to 'notice' the non-occurrence of a particular meaning, it is necessary to hear a scrambled sentence in a situation in which this would be true but the non-scrambled utterance would not be. As we have seen, however, whenever the scrambled utterance is true, the non-scrambled

---

(regardless of the alternation in (i)). This would predict targetlike behaviour from very early on. As we shall see in Chapter 6, however, this is not consistent with the results obtained in the present study. The second and most important reason why I consider it unlikely that the interpretive constraint on the double object sequence illustrated in (i)-b will have the proposed facilitatory effect is because this requires the L2er to somehow equate the scopal relations between the two arguments of a ditransitive verb with those between the single argument of a transitive verb and a sentential scopal operator such as negation or a frequency adverbial. It is unclear why L2ers should do this, especially given that in English, direct objects and sentential scopal operators such as negation have different scopal relations. Note, furthermore, that it is sometimes possible for a universally quantified direct object to take scope over an existentially quantified indirect object. This is illustrated with an example from Su (2001:168, ex. 29) in (ii).

(ii) Generally, I give a tourist every leaflet

[28] The following books, which include most of the standard textbooks, were consulted: *Dutch: An Essential Grammar* (Schetter and van der Cruysse 2002), *Teach Yourself Dutch Grammar* (Quist and Stuik 2000), *Essential Dutch Grammar* (Stern 1984), *Grammatica in Gebruik. Nederlands voor Anderstaligen* (Bakx, Jetten and Korebits 1999), *Help! Een Cursus voor Buitenlanders* (Ham, Tersteeg and Zijlmans 1998), *Nederlands voor Buitenlanders: De Delftse Methode* (Montens and Sciarone 1991) and *Code Nederlands* (van Kalsbeek, Huizinga and Kuiken 1998).

[29] In other words, indirect negative evidence (Chomsky 1981) would be required. According to Philip (2003), the L1 acquisition of the comprehension of scrambled indefinite objects proceeds on the basis of indirect negative evidence and this is why it is so slow and gradual. As noted above, without a clear model of how indirect negative evidence actually works, this proposal remains difficult to assess (see §3.3.1.2 and discussion there). In Chapter 6, I nevertheless consider whether the L2 data are compatible with such an approach.

utterance is also true. In other words, for the contrast given in (35), there is no context which could dissuade the English-speaking learner of Dutch that (35)-a is not simply a rewrite of (35)-b.

The situation for scrambling across negation is slightly different, however.[30] There is one instance in which the scrambled utterance, illustrated in (36)-a, is true and the non-scrambled utterance, given in (36)-b, false, and this could, in principle, trigger the acquisition of the relevant constraint. I will show, however, that the relevant circumstances are likely to be infrequent and very often not unambiguous.

(36) a. De jongen heeft een vis niet gevangen
the boy has a fish not caught
'The boy didn't catch a (certain) fish.'

b. De jongen heeft geen vis gevangen
the boy has no fish caught
'The boy didn't catch a(ny) fish.'

The scrambled and non-scrambled utterances given in (36) are both true in a situation where no object (out of any number of objects) is manipulated. As soon as there are two or more objects, however, and one or more objects are manipulated and exactly one object is not manipulated, the scrambled utterance remains true but the non-scrambled utterance is false. This is the situation depicted in Krämer's experiment discussed in §3.3.1.1, that is, where there are three fish in the pond, two of which are caught, leaving one fish remaining, which is not caught. If L2ers were to induce the constraints on scrambling from the target language input they hear, such circumstances would be crucial.

Let us consider some possible scenarios in which acquisition might take place on the basis of such situations. Imagine a situation where an L2er attempts to convey the message that out of three available cookies, there was one that Bob did not eat. If this learner has not yet acquired scrambling, she might use an utterance such as in (37)-a, with the suppletive form *geen*, or (37)-b, if she has not acquired *geen*, or alternatively, (37)-c, with the stressed indefinite determiner.

(37) a. Bob heeft geen koekje gegeten
Bob has no cookies eaten
'Bob hasn't eaten a cookie.'

b. Bob heeft niet een koekje gegeten
Bob has not a cookie eaten
'Bob hasn't eaten a cookie.'

[30] This follows from the different entailment relations which obtain between scrambled and non-scrambled orders in negative sentences and in sentences with *twee keer* (cf. §2.1.2).

c. Bob heeft niet één koekje gegeten
Bob has not a cookie eaten
'Bob hasn't eaten a cookie.'

In such circumstances, the utterance in (37)-a is false and this might result in the L2er's interlocutor responding with something like (38) or (39):

(38) Hij heeft wel een koekje gegeten. Twee zelfs!
he has indeed a cookie eaten two even
'He has eaten a cookie. Two even!'

(39) a. Nou, hij heeft één koekje niet gegeten
well he has one cookie not eaten
'Well, he hasn't eaten one (certain) cookie.'

b. Nou, hij heeft er ééntje niet gegeten
well he has one not eaten
'Well, he hasn't eaten one of them.'

The response in (38) would indicate to the L2er that her utterance did not convey the intended meaning, but because it does not contain a scrambled object, it would not provide any evidence of how that meaning should be conveyed in the target language. Both responses in (39) *do* contain a scrambled object but that object contains the stressed determiner/cardinal *één* 'one', and hence, they are ambiguous with respect to how the intended meaning should be conveyed: at issue could be the position of the object (scrambled) or the form of the object (*één koekje* rather than *(g)een koekje*) or both. If the L2er were to produce the utterance in (37)-b, the native-speaker interlocutor may interpret (37)-b as (37)-a, which as we have seen, would not necessarily result in an interaction which would provide the L2er with the relevant information to acquire the target <form, meaning> pairing. Alternatively, native-speaker interlocutors might interpret it as expressing constituent negation if the object were stressed (i.e. Bob didn't eat a cookie, he ate something else), and they might subsequently question this interpretation (e.g. *Oh no?! What did he eat then?*), but it seems unlikely that in such an interaction, they would produce an utterance containing a scrambled object (not realised as *één*). If the L2er were to produce a non-scrambled utterance as in (37)-c, with the stressed determiner/ cardinal *één* 'one', this is also unlikely to elicit an informative response from the native-speaker. This is because (37)-c is actually true in the scenario outlined above: it is true that Bob did not eat one cookie, because he ate two.

The problem of *één* versus *een* is irrelevant when another numeral determiner is used. Could utterances such as (40)-a provide the necessary trigger to acquire the interpretive constraint on scrambling, as Krämer (2000:179) suggests?

(40)	a.	Bob	heeft	twee koekjes	niet	gegeten
		Bob	has	two cookies	not	eaten
		'Bob hasn't eaten two cookies.'				
	b.	Bob	heeft	niet	twee koekjes	gegeten
		Bob	has	not	two cookies	eaten
		'Bob hasn't eaten two cookies.'				

In order for this to be the case, there must be a situation in which (40)-a is true and (40)-b is false. Returning to our scenario with three cookies; imagine this time that Bob eats just one, so that there are two cookies which he does not eat. In such a situation, (40)-a is true but (40)-b is *also* true, because Bob did not eat two cookies, he ate one. This holds for any situation in which the number of cookies Bob did eat is different from the number of cookies which he did not eat. In other words, utterances such as (40)-a can only be informative with respect to the interpretive properties of scrambling in a situation where the number of unmanipulated objects is equal to the number of manipulated objects, that is, for (40), where Bob ate two cookies and left two cookies. It is certainly questionable whether (i) all L2ers would encounter such scenario-utterance pairings and (ii) whether any interaction surrounding such utterances would be sufficient to make this difference in meaning clear. This problem is compounded by the fact that it is in precisely such circumstances that the English equivalent, given in the glosses in (40), is ambiguous, so even if the input could potentially help the L2er, their L1 would be pulling them in an alternative direction.[31]

The preceding discussion has demonstrated that it is very improbable that knowledge of the interpretive constraints on scrambled indefinite objects can be gleaned from the input. Scrambled indefinite objects are quite infrequent in the first place and the circumstances in which they would have to occur in order to be informative for the language learner are (a) even more infrequent and (b) unlikely to be available to all learners.[32] The L2 acquisition of the interpretive restrictions on scrambled objects by native speakers of English therefore constitutes a poverty-of-the-stimulus problem. If there exist L2ers who are able to come to know these restrictions, it must be concluded that their interlanguage grammars are constrained by UG. The claim that the

---

[31] It is of course logically possible that, say, a native speaker may explicitly inform the L2er that her non-scrambled utterance is incorrect and that if, for example, she wants to say that there is one cookie which was not eaten she should use the scrambled form. Given that native speakers are usually not consciously aware of the interpretive difference between the scrambled and non-scrambled forms, this also seems highly unlikely.

[32] To the best of my knowledge, there are no studies which present frequency data in Dutch dealing specifically with scrambled indefinite objects. Hence, no figures are provided here. Most native-speaker linguists would agree, however, that indefinite objects occur in scrambled position significantly less frequently than in non-scrambled position.

acquisition of the interpretive constraints on scrambling constitutes a poverty-of-the-stimulus problem also holds for L1 acquisition. What exactly the relevant UG constraints might look like, that is, how the interpretive constraints on scrambled indefinites might be acquired, is addressed in Chapter 7 (§7.4).

One possible objection to the approach adopted here concerns its applicability to child L1 acquisition. To state that the L2 acquisition of property X requires UG implies that the L1 acquisition of this property is also constrained by UG. The studies by Krämer (2000) and Philip (2003) suggest that targetlike comprehension may not be in place until very late and this is unexpected for a property of language whose acquisition is constrained by UG. I believe that the empirical observations made in the aforementioned studies – which are also replicated here (in Chapter 6) for L1 children – are not irreconcilable with the claim that the acquisition of scrambling is in a certain sense constrained by UG. I will suggest, on the basis of new child L1 production data (in Chapter 5), that target-language knowledge of scrambling must be in place at an earlier age and that consequently the delay observed in the comprehension data results from discourse/pragmatic factors but these discourse/pragmatic cannot be the sole driving force behind the acquisition of the interpretive constraints on scrambling.

Thus far, the discussion in this section has concentrated on scrambled indefinite object NPs. This thesis also examines the acquisition of scrambled definite object NPs across negation. As noted in Chapter 2 (cf. §2.1.1, §2.4, §2.5.2), negation and focus interact such that negative sentences containing scrambled definite NP objects, exemplified in (41)-a, are generally interpreted as expressing sentential negation, whereas those with non-scrambled definite NP objects, exemplified in (41)-b, are assigned a contrastive negation interpretation. This is for the following reason. For the object to have a contrastive reading, negation must be associated with the object only. On the assumption that negation associates with focus, this would require only the object to be in focus but this is only possible when it occurs in non-scrambled position. When the object is in the scrambled position, it cannot be the only focus, and therefore, (with neutral stress) a contrastive negation reading is not possible. Note, however, that there are no truth-conditional differences between (41)-a and (41)-b. (SMALL CAPITALS indicate main sentential stress and not contrastive stress.)

(41) a. De jongen heeft de vis niet geVANgen
the boy has the fish not caught
'The boy didn't catch the fish.'

b. De jongen heeft niet de VIS gevangen
the boy has not the fish caught
'The boy didn't catch the FISH. (He caught something else)'

The interaction between focus, negation and scrambling is not the subject of classroom instruction. Furthermore, the L2ers' L1, English, does not have the option of scrambling and, consequently, the difference between sentential and contrastive negation is achieved using stress-shift, as illustrated in (42). (Again, CAPITALS indicate main sentential stress and not contrastive stress.)

(42) a. The boy didn't catch a FISH
b. The boy didn't CATCH a fish
c. The boy did NOT catch a fish

The sentence in (42)-a, which has neutral sentence stress, could be interpreted as contrastive negation on the object or as sentential negation. In order to ensure that a contrastive negation interpretation of the object is ruled out, the stress is shifted to the verb, as in (42)-b, or to the negator, as in (42)-c (cf. §2.5.2).

Above, I argued that the L2 acquisition of scrambled indefinite objects by native speakers of English constitutes a poverty-of-the-stimulus problem. The L2 acquisition of scrambled definite objects across negation presents a comparable but not identical situation. The acquisitional task faced by the English-speaking L2er of Dutch is to come to know that when sentential negation is intended, the scrambled order should be used rather than stress-shift. How exactly L2ers should come to know this is far from obvious. This is because if L2ers transfer stress-shift from their L1 and apply this to their interlanguage production, this will result in utterances which, although not targetlike in the sense that they do not employ scrambling, are nonetheless grammatical. For example, L2ers might produce (43)-a, where the stress has been shifted from the most deeply embedded constituent, the definite NP object, to the verb, or (43)-b, where the stress has shifted to the negator.

(43) a. De jongen heeft niet de vis geVANgen
the boy has not the fish caught
'The boy didn't CATCH the fish.'

b. De jongen heeft NIET de vis gevangen
the boy has not the fish caught
'The boy did NOT catch the fish'

Both of these utterances are not ungrammatical, but native speakers prefer the scrambled variant in (41)-a. The sentence in (43)-a is likely to be interpreted as expressing a contrast on the verb (as in 'the boy didn't *catch* a fish, he did something else to it'), although if this were the intended interpretation, native-speaker informants suggest they would scramble rather than produce (43)-a (as predicted by Reinhart 1995 and Neeleman and Reinhart 1998; cf. §2.3.2.2). The sentence in (43)-b would be acceptable as an expression of sentential negation, but once again, it is dispreferred in comparison to the scrambled sentence in

(41)-a. This is especially so when the definite NP object is anaphoric because anaphoric definite NPs scramble more often than not (de Hoop 2003).

It is difficult to imagine a situation in which the production of (43)-a or (43)-b might lead to a communicative misunderstanding such that the preference for the scrambled form would be transparent. Furthermore, the TL input will not provide any evidence that using stress-shift rather than scrambling is non-targetlike because stress-shift is also possible in Dutch. In addition, definite NP objects generally scramble optionally, which would render the target language input even less clear. All in all, then, although it is difficult to speak of a poverty-of-the-stimulus problem as standardly assumed, the sources available to the L2er, namely the L1, the TL input and any classroom instruction, are unlikely to lead to targetlike knowledge of the difference between scrambled and non-scrambled definite NP objects in negative sentences. Thus, if English-speaking L2ers can nevertheless come to know what the target language preferences are, their interlanguage grammars must be constrained in the same way as the native speaker's grammar, that is, the option of stress-shift will be dispreferred because the grammar offers another (cheaper) option, scrambling.

To summarise: In this section, I argued that the acquisition of direct object scrambling in Dutch constitutes a poverty-of-the-stimulus problem. English-speaking L2ers do not receive classroom instruction on this topic, they cannot induce the interpretive restrictions on scrambling from the input, nor can they transfer them from their L1. Consequently, if they come to know these restrictions, their interlanguage grammars must be constrained in the same way as native speaker grammars.

The final section of this chapter presents the predictions which will be tested in the remainder of the thesis.

3.5 General predictions

The central question addressed in this thesis concerns the extent to which child L2 acquisition is like adult L2 acquisition or child L1 acquisition (or both, or neither). In order to answer this question, English-speaking child L2ers will be compared with, on the one hand, English-speaking adult L2ers, and on the other, with L1 Dutch children, in their acquisition of one particular property of the target language, namely direct object scrambling in Dutch. This section outlines the general predictions for these two comparisons. These will be fleshed out in Chapters 5 and 6 in terms of the specifics of the experimental tasks which are employed. These predictions are based on (i) the approach proposed by Schwartz (1992; 2003; 2004) outlined in §1.2.1.2, (ii) previous findings on child L2 acquisition detailed in Chapter 1, (iii) previous findings on the L1 and (adult) L2 acquisition of scrambling presented in this chapter, and (iv) the particular acquisitional task which scrambling in Dutch presents for the English-speaking L2er, as outlined in the preceding section.

3.5.1 Child L2 ~ Adult L2

The rationale behind the child L2 ~ adult L2 comparison carried out here is based on Schwartz (1992; 2003; 2004). This approach, which is designed to arbitrate between UG-based and alternative, general learning approaches to adult L2 acquisition, is schematised as follows:

(44) a. Child L2 development = Adult L2 development:[a,b]
→ Adult L2 acquisition is constrained by UG
[a] *holding the L1 constant*
[b] *for UG-governed language properties*

b. Child L2 development ≠ Adult L2 development:[a,b]
→ Adult L2 acquisition is not constrained by UG
[a] *holding the L1 constant*
[b] *for UG-governed language properties*

(Schwartz 2004:39)

Schwartz argues that assuming (i) that child L2 acquisition is driven by UG and (ii) that because (L1 and L2) children make use of UG (domain-specific principles), the general learning principles in question are more relevant to adult L2ers, then (iii) comparing developmental sequences of child L2ers with those of adult L2ers, while holding the L1 constant, will provide evidence for or against UG involvement in adult L2 acquisition. This is schematised in (44) (see §1.2.1.2 for more details).

Schwartz's approach juxtaposes UG-constrained L2 development against a domain-general approach. Let us for a moment consider what such general learning principles might look like. Meisel (1997:258) suggests that L2ers, 'rather than using structure-dependent operations constrained by UG, resort to linear sequencing strategies which apply to surface strings'. On such an approach, L2ers must rely on surface word order only. How and whether they match the word order patterns they hear to the meaning which these patterns have is not clear. However, assuming for argument's sake, that such pattern~meaning matching is possible, this would mean that if the English-speaking L2ers under investigation here come to know that scrambled indefinite objects cannot have a non-specific interpretation, they must have done so on the basis of the input and in particular, on Meisel's approach, solely on the basis of the linear word order they encounter in the input.

In §3.4, it was shown that the only circumstance in which the input might provide information about the target constraint is when a sentence containing an indefinite object scrambled across negation occurs in a situation where one object from a set of objects has been manipulated in some way and there are one or more objects which have not undergone such a manipulation. On (the best possible interpretation of) such a pattern-matching approach, it would be crucial for L2ers to take the following steps. They would have to (i) encounter such pattern~meaning pairings, (ii) encounter them frequently enough for the Object-Negation-Verb sequence to be considered some kind of pattern, (iii) notice that another pattern, namely Negation-Object-Verb, is also possible, (iv) relate these two patterns to each other, (v) notice (and presumably at some level remember) the meaning which the Object-Negation-Verb patterns have, and (vi) realise somehow that if the Negation-Object-Verb pattern had occurred with that meaning, it would be false.

It is questionable whether, based on linear word order alone, L2ers would be able to make many of these steps. First, as noted above, it is not clear whether such pattern~meaning matching is possible on an approach such as Meisel's. Second, if interlanguage grammars operate on surface strings only, it is not evident on what basis L2ers should link the Object-Negation-Verb order to the Negation-Object-Verb order; this is a crucial prerequisite for the final steps in the hypothesised learning process to take place. Why should the Object-Negation-Verb sequence not be related to a sequence in which the object is topicalised, for example, or to Adverb-Negation-Verb sequences, which exhibit the same linear order but with different constituents? Assuming for the moment that, despite the problems listed above, L2ers would somehow be able to use linear sequence strategies to acquire the interpretive constraints on scrambled indefinites, given that the only relevant data in the input concern scrambling across negation, it is not clear how and whether, without any knowledge of the underlying structure, L2ers would be able to extend this knowledge to scrambling across *twee keer* or any other adverbial, which they would have to do given that the input is uninformative with respect to scrambling across *twee keer*, for example. All in all, then, without extensive modification, a general learning principle such as pattern-matching would not allow the L2er to acquire targetlike knowledge of the interpretive constraints on scrambling.

In the previous section, it was argued that precisely because the relevant information is not available to the English-speaking L2er, either in the TL input, from L1 transfer or from classroom instruction, the acquisition of the interpretive constraints on scrambling by English native speakers constitutes a poverty-of-the-stimulus problem, and if L2ers develop targetlike knowledge of these constraints, their interlanguage grammars must be constrained by UG. The comprehension studies reviewed earlier in this chapter showed that L1 children failed to interpret scrambled objects in a targetlike fashion until very late, indicating that the acquisition of scrambling cannot solely depend on UG, however. Rather, some aspect of discourse/pragmatics must also be involved. This rather complicates the logic of the child L2 ~ adult L2 comparison as set out above. Such a comparison remains interesting nevertheless, because it may help us tease apart the role played by such discourse/pragmatic factors in the (L1 and L2) acquisition of scrambling.

With respect to the child L2 ~ adult L2 comparison, the following predictions are tested. Recall that these are based on (i) the assumption that child L2 acquisition is constrained by UG and (ii) the hypothesis that adult L2 acquisition is also constrained by UG.

PREDICTION ❶. CHILD L2 ~ ADULT L2: DEVELOPMENTAL SEQUENCES
L2 children and L2 adults will pass through the same developmental sequence.

PREDICTION ❷. CHILD L2 ~ ADULT L2: DISCOURSE/PRAGMATIC FACTORS
a. L2 children within the same age range as L1 children with limited discourse integration will differ from L2 adults in their interpretation of scrambled indefinite objects.
b. L2 children outside of this age range should pattern similarly to L2 adults.

We now turn to the predictions for the comparison between L2 children and L2 adults.

3.5.2 Child L2 ~ Child L1

In Chapter 1, I argued that child L2 acquisition may provide an interesting testing ground for accounts of L1 acquisition which make recourse to age-related factors. Although Krämer's (2000) account of the L1 acquisition of scrambling does not invoke maturation in a strict sense (as in, for example, Rizzi 1993/1994, on Root Infinitives), it is age-related in that it relies on some other factor, namely discourse integration, which is thought to cause L1 children problems until a relatively late age, and as such, the logic of the child L2 ~ child L1 comparison is applicable here. This logic is as follows: L2 children who are younger than the age at which an age-related developmental stage purportedly occurs in L1 acquisition are not expected to be in that stage in their L2 development (although they may of course pass through this stage later, once they reach the relevant maturational point).

With respect to the child L2 ~ child L1 comparison, the following predictions are therefore tested:

PREDICTION ❸. CHILD L2 ~ CHILD L1: DEVELOPMENTAL SEQUENCES
As a result of L1 transfer, L2 children will pass through a different developmental sequence from L1 children.

PREDICTION ❹. CHILD L2 ~ CHILD L1: DISCOURSE/PRAGMATIC FACTORS
L2 children within the same age range as L1 children with limited discourse integration will behave similarly to L1 children in their interpretation of scrambled indefinite objects.

These predictions will be made more specific within the context of the experiments in Chapters 5 and 6.

3.6 Summary

To come to know the interpretive constraints on scrambling requires sophisticated knowledge of the interaction between syntax, semantics and discourse/pragmatics. By examining the development of this property of Dutch in three different groups of learners, this thesis endeavours to shed light on the role played by these knowledge types in the acquisition process. The review of previous research on the acquisition of scrambling showed that in production, both L1 children and L2 adults pass through a developmental stage where scrambling is either absent or optional. Targetlike scrambling by L1 Dutch children – at least of definite NP objects – was found to be in place at around age three years. This contrasts with the results of studies on the comprehension of scrambling, which showed that L1 children assign a non-targetlike interpretation to scrambled indefinite objects up to age seven years and older. The L2 acquisition of the interpretive constraints on scrambling has not yet been systematically investigated and little, if anything, is known about the acquisition of scrambling by L2 children. The present study aims to fill this gap.

A brief discussion of some of the literature on the syntax-semantics interface in (adult) L2 acquisition highlighted the task faced by the English-speaking L2ers in the present study. It was claimed that in order to acquire the interpretive constraints on scrambling in Dutch, these L2ers must overcome the poverty-of-the-stimulus: the fact that the scrambled form cannot be associated with a non-specific interpretation cannot be induced from the TL input, it cannot be transferred from the L1, and is unavailable in the L2 classroom.

On the basis of the literature reviewed in this chapter and the approach outlined in Chapter 1, certain similarities and differences were predicted for the two comparisons carried out here. On the assumption that L1 transfer will affect both L2 groups and if adult L2 acquisition is constrained by UG, it was predicted that L2 children will pass through the same developmental sequence as L2 adults and this sequence will differ from that of L1 children. The comparison between L2 children and L2 adults in their interpretation of scrambled objects is complicated by discourse/pragmatic factors. It was predicted that those L2 children who are in the same age range as the non-targetlike L1 children should behave similarly to the L1 children and that both these groups would pattern differently from the L2 adults. These predictions will be tested by carrying out the same production and comprehension experiments with each of these three groups. Before moving to the experimental data, however, it is first necessary to outline the proficiency measure which will form an integral part of the child L2 ~ adult L2 comparison.

CHAPTER 4

MEASURING L2 PROFICIENCY

Introduction

This thesis examines the L2 acquisition of direct object scrambling in Dutch by English-speaking children and adults. In order to be able to make cross-group comparisons in this domain, it is necessary to have some independent means of matching and comparing subjects with each other. This chapter presents the language proficiency measure which will be used for this purpose.

In L2 acquisition research, the construct 'language proficiency' is defined in numerous ways. As well as being used as a global indicator of an L2er's abilities in the target language, it is also used to refer to specific aspects of linguistic competence, such as phonological, syntactic, morphological, lexical and/or discourse skills. In general, language proficiency is divided up into knowledge and some aspect of use (control/ communicative competence) and it often involves one or more of the following dichotomies: productive vs. receptive, written vs. oral, communicative vs. grammatical, etc.

Defining proficiency – one of the goals of a whole field of applied linguistics (see e.g. papers in De Jong and Verhoeven 1992) – is well beyond the scope of this thesis. Nevertheless, in order to be able to proceed, a concrete definition of this construct is required. Intuitively, proficiency might be defined as 'a person's overall competence and ability to perform in L2' (Thomas 1994:330, fn. 1). Such intuitive definitions, however, remain somewhat tautological (Verhoeven and Vermeer 1989:26). For the purposes of the present study, the construct of L2 proficiency is operationalised as 'the ability to produce lexically, morphologically and syntactically complex and accurate utterances in the target language (TL)'.[1, 2]

This definition focuses on knowledge, that is, linguistic competence, rather than use/control (although when testing knowledge some aspect of use/control will inevitably be involved) and hence it excludes pragmatic skills, many of which would be classified as properties of language use. An additional reason for excluding pragmatic skills is that operations above the sentence level are an important part of scrambling and it is (part of) this aspect of language

---

1 This is not to say that other aspects of language, such as phonology, should not be considered part of proficiency. Rather, this is simply excluded for the purposes of the present study. This is not unusual. For example, cloze tests, which are regularly used to assess (adult) L2 proficiency (see below) do not assess phonology.

2 The proficiency measure is based on data collected using a production task but the assumption is that this task taps the L2ers' underlying linguistic competence.

which has been claimed to be in part responsible for the acquisition of scrambling (cf. Chapter 3). Including such skills in the proficiency measure would therefore make it less independent.

This chapter is organised as follows. The following section (§4.1) presents the goals of the proficiency measure used here. Section 4.2 briefly sketches some of the ways in which proficiency has been measured in L2 research and it introduces the picture description task which is employed here. The data collected using this task are used to derive a proficiency score which is based on three component parts measuring (i) morphosyntactic complexity, (ii) lexical complexity and (iii) morphosyntactic and lexical accuracy. An overview is given in §4.3; this provides the reader with sufficient information to be able to progress directly to the experimental work on scrambling presented in Chapters 5 and 6, if so desired. The remainder of the chapter consists of a detailed motivation for the particular measures used in the proficiency score. There are several reasons why this is provided. First and foremost, the proficiency score plays a crucial role in the analyses presented in the remainder of this thesis. It is therefore imperative that the contents of this score are made explicit. Second, a detailed examination of several alternative measures highlights some of the methodological issues inherent to a child L2 ~ adult L2 comparison. Third, with respect to the use of MLU as a measure of linguistic development, the review in §4.4.1 brings together the literature on younger and older L1 children and L2 adults in one place. The bulk of the chapter (§4.4) involves a critical evaluation of this and various other measures of complexity, including the application of such measures to data from L1 children and adults specifically collected for this purpose. A detailed explication of the accuracy measure is provided in §4.6. The final section (§4.7) outlines how the different parts of the proficiency measure are combined using 'principal components analysis' to obtain one final proficiency score per subject.

4.1 The goals of a proficiency measure in the present study

Thomas (1994) notes that in L2 research, proficiency measures are necessary when (i) the researcher wishes to compare different groups of L2ers in their acquisition of a given phenomenon and (ii) when cross-sectional data are used to derive developmental sequences. These are both goals of the present study: (i) child L2ers will be compared with adult L2ers and (ii) cross-sectional data from L2ers with different proficiency levels will be used to determine the stages L2ers pass through in their acquisition of direct object scrambling. In order to achieve these goals, we therefore need an approximate indication of the subjects' proficiency level in Dutch which is independent of the main experimental data.

As noted above, here, proficiency is operationalised as 'the ability to produce lexically, morphologically and syntactically complex and accurate utterances in the TL'. Phonology is excluded. The proficiency measure therefore incorporates three aspects of linguistic ability (syntax, morphology, lexicon) measured in terms of their complexity and accuracy. This measure is not intended to be a measure of global proficiency, because, for reasons noted below in §4.2.1, it does not include pronunciation/phonology, discourse/pragmatic skills or comprehension. Rather, it is intended to provide an approximate but nevertheless reliable indication of L2ers' differing proficiency levels such that the goals outlined above may be achieved.

The next section provides a brief sketch of the types of proficiency measures which can be found in the L2 literature, before outlining the method which will be used.

4.2 How to measure proficiency

Deciding how proficiency should be measured involves deciding which type of task to use to collect the necessary data and selecting relevant variables to include.

4.2.1 Choosing a task

There were several practical considerations in deciding what type of task to use to measure proficiency. Firstly, the task could not involve reading/writing, because this would have been too difficult and even impossible for some of the subjects, and in particular, for the child L2ers. Consequently, a traditional cloze test, often used as a proficiency measure for adult L2ers, was immediately ruled out. Secondly, the task should not be too 'test-like', in the sense that it should not resemble the type of language tests commonly administered in L2 classrooms. This is to avoid the application of metalinguistic knowledge and rules, something which is considered to be different from L2 proficiency as defined here. Given that metalinguistic knowledge develops with cognitive maturity, this was considered more likely to be an issue with the older child L2ers and adult L2ers (Appel 1984:139). Hence, the use of such a test could introduce an unwanted variable into the proficiency scores, which would at least partly co-vary with age (cf. §1.4.1.2). This is not to say of course that metalinguistic knowledge is not used in more naturalistic tasks (for example, by 'monitoring' or applying rules online, etc.) or that this does not vary across speakers, but more naturalistic tasks at least do not *encourage* the use of such knowledge. One final practical consideration in deciding which task to use was time restrictions. Due to the limited time available with each subject, especially the child L2ers, who were exempted from their regular classes in order to participate, the proficiency measure could take no longer than 15 minutes.

In a survey of 157 articles taken from four L2 acquisition journals,[3] Thomas (1994) observes that L2 proficiency tends to be measured in four different ways: (i) impressionistic judgement, (ii) institutional status, (iii) in-house assessment instrument and (iv) standardised test. Each of these will be addressed in turn.

Impressionistic judgement, which involves 'asserting that a learner has a given level of control over [the] L2, on the basis of the experimenter's unsupported evaluation, or the evaluation of some other (often unspecified) person' (Thomas 1994), has the clear disadvantage of lacking generalisability: one person's 'advanced' is not another's. As Thomas (1994:317) points out, institutional status, for example, first-year students vs. final-year students, suffers from a similar weakness: standards are determined in different ways in different institutions, and hence final-year students at one university might not be the same as final-year students at another university. Furthermore, in the pool of subjects examined here, there is no such common denominator with which they could be compared: the child L2ers attended several different schools and the adults had followed different courses.

Thomas (1994) reports that in-house assessment instruments come in two flavours: (i) in-house placement instruments used to assess students at the beginning or during foreign language programmes and (ii) tests developed privately by researchers, but not with the explicit purpose of assigning students to pre-defined levels within an institution. It is this latter category which is closest to the type of proficiency measure developed here. Thomas (1994:322) writes that both these types of in-house assessment instruments have the advantage of ensuring that all subjects are tested in a uniform fashion. This means that any resulting proficiency scores are internally consistent within the sample and 'subgroups may be compared with respect to proficiency on some rational basis'.

There are, however, a number of disadvantages to this type of proficiency measure, as Thomas notes. Firstly, the same criticism levelled at impressionistic judgement and institutional status, namely a lack of generalisability, is also valid here. Such tests are not standardised and any conclusions based upon them are restricted. As noted above, however, they do test all subjects in a uniform fashion. Given the lack of any other independent variable arguably related to proficiency common to the two main groups tested here (adult/child L2ers), this is crucial. The problem of lack of generalisability is heightened when researchers fail to explain what exactly such measures involve. One way we can at least attempt to reduce this problem is to make the contents of a given measure explicit. In this way, other researchers will be able to make informed judgements about the extent to which two or more proficiency measures are comparable and furthermore, if desired, exact

[3] The journals were: *Applied Linguistics*, *Language Learning*, *Second Language Research* and *Studies in Second Language Acquisition*. All volumes from 1988 until 1992 were surveyed.

replication will also be possible. A second disadvantage of privately developed proficiency measures is that they may vary with respect to their validity and reliability.[4] Providing independent motivation for the decisions made about what to include or exclude in such a privately developed proficiency measure is one way in which the validity and reliability of a particular measure can be improved. This will be the approach adopted here. A third problem observed by Thomas was that many researchers in her review did not explicitly state whether and to what extent their proficiency measure overlapped with the content of their experimental materials. As mentioned above, the proficiency measure used in the present study is intended *not* to overlap with the experimental materials (on scrambling), that is, knowledge of scrambling (in its narrowest sense) is not part of the knowledge measured in the proficiency score. This is because, as stated in the previous section, one of the objectives of the proficiency measure is to provide an *independent* indication of the learners' proficiency (as defined above) so that we can use this to compare subjects across different populations (child vs. adult) and across different levels (low vs. mid vs. high proficiency) with respect to their knowledge of scrambling. It is also for this reason that pragmatic skills and utterances including (non-)scrambling were excluded from the data used to derive the proficiency score.[5,6]

The final proficiency measure in Thomas' list is the standardised test. There are several standardised tests available for Dutch. These include the TAK test (*Toets Allochtone Kinderen* 'Immigrant children test'; Verhoeven and Vermeer 1989), the ISK tests (*Internationale Schakel-Klassen* 'International transition classes' from CITO (Dutch testing agency)) and the NIVOR test (*Niveauvorderingentoetsen* 'Level Assessment tests', also from CITO). While each of these tests has the advantage of being standardised, some are predominantly or exclusively in written form. Also, they are each designed for different age groups: the TAK is designed for children, the ISK for adolescents and the

[4] Assessing the validity of a given measure involves determining whether it actually measures the construct it claims to measure. In assessing its reliability, we ask whether it does this consistently (that is, whether the method is accurate).

[5] Here, one might object that knowledge of scrambling also involves knowledge of syntax, yet a syntactic component is still included in the proficiency measure. Rather than exclude syntactic aspects of language from the proficiency measure altogether, I simply excluded any utterances involving (non-) scrambling (see §4.6.3). Given that the discourse/pragmatic skills (discourse integration, Schaeffer's pragmatic principle) which have been claimed to be involved in scrambling are more general, it is not clear how these could be factored out of any discourse or pragmatic component in the proficiency measure.

[6] There is no reason to believe that by using measures of syntactic accuracy and complexity in the proficiency score, this score should be measuring exactly the same thing as the scrambling experiment. As we shall see, the syntactic knowledge included in this score encompasses much more than scrambling and (non-)scrambled utterances are excluded. The expectation is that just as knowledge of scrambling will develop and become progressively more nativelike, so will these and other properties of the interlanguage grammar.

NIVOR for adults. This is problematic given the purpose of the proficiency measure here is to find a way of comparing these groups. Furthermore, the TAK was developed with specific L1 groups in mind, that is, Moroccan Arabic, Turkish and Surinamese primary school-aged children. This means that for present purposes, some modification would be necessary, either to make the tests suitable for English-speakers and/or for different age groups, which would, of course, undermine the advantage of external validation which standardisation confers upon them. Furthermore, these tasks often focus very clearly on the application of rules, for example, how to form noun plurals, how to inflect verbs, which is something which we want to avoid as this could encourage the use of metalinguistic knowledge (especially on the part of the adults). One final problem with such tasks is that they take too long to administer, given the time limitations in the present study. Given such considerations, using one of these standardised tests was considered unsuitable.[7]

Instead, a privately developed proficiency measure was opted for. Specifically, following Whong-Barr and Schwartz (2002), I decided to use a picture description task to elicit (semi-)spontaneous data. There are several advantages to such a task. Firstly, it is based on spoken language only and is therefore suitable for use with the younger subjects. Secondly, it involves an activity, namely describing pictures/telling a story, which most (if not all) subjects are familiar with. Thirdly, it requires the subject to focus on content rather than on form, which consequently, also serves to help the subject relax. Fourthly, it can be worked into the test procedure quite easily and it can also be used as a sort of distracter to split up parts of the experimental procedure. Furthermore, it is not very time-consuming (10-15 minutes maximum).

Before outlining how the data obtained using this task will be analysed, the data collection procedure and task are outlined in more detail.

[7] One possible solution to this latter problem would be to use a sub-component of the *TAK* test only. This is the approach adopted by Blom and Polišenskà (to appear). They use one part of the *TAK* test, a sentence imitation task, in their comparative study of L2 children and adults (cf. 1.4.3.4). Their subjects were (mainly) speakers of Moroccan Arabic and Turkish and hence, the test, which was designed for these languages, is certainly suitable in this regard. However, as noted above, the *TAK* was designed for and validated using children. It is not clear to what extent its use can be extended to L2 adults, particularly given that children and adults have different working memory capacities and that working memory is an important factor in an imitation task. In this light, it is possible that this task is not suitable as a means of comparing child and adult subjects.

4.2.2 Picture description task: Procedure

The materials used in this task involved sets of between four and eight pictures which depicted a series of events.[8] Some of the events involved children's characters, such as Paddington Bear and Donald Duck. When subjects did not know these characters, they were provided with the name or they sometimes spontaneously created their own name. All the actions depicted in the pictures were considered general enough so as not to (dis)favour either the child subjects or the adult subjects. For example, they included planting flowers and watering them, digging a hole in the beach, baking cookies, having an accident on a bicycle, making a puppet show, painting a picture and a cat climbing on a roof. Example pictures are given in Appendix B. There were two different sets of pictures, one used as part of the production experiment reported on in Chapter 5 and the other as part of the comprehension experiment detailed in Chapter 6. This was because several subjects took part in both experiments.

At the appropriate point in the experimental procedure, subjects were presented with these pictures and asked to describe/tell a story about what they saw. During the task, subjects were encouraged to speak as much as possible, and when necessary, the experimenter(s) provided prompts and asked questions designed to elicit more data, such as '*And what happened next?*'. I tried to achieve a balance between putting the subjects at ease so that they could perform the task to the best of their ability and not interfering so as to ensure that the data obtained for each subject were as uniform as possible. When requested to by the subjects, the experimenter(s) provided any target-language words they did not know, but in general, the experimenter(s) avoided interacting too much with the subjects as this encouraged the use of elliptical utterances (see §4.3.1 for relevant discussion).

The task was recorded using a Tascam DA-P1 DAT recorder and Crown PZM185 microphone and the data were subsequently transcribed in CHAT format using the CLAN programme available via CHILDES (MacWhinney 2000). Transcriptions were mostly carried out by one of the experimenters who was present at the time of recording (either the present author or a native-speaker student assistant) and they were checked by either the other experimenter or another native-speaker student assistant. A handful of recordings were also transcribed by a native speaker who was not present at the time of recording. He was given a copy of the pictures used in the task to provide context and his transcriptions were checked by one of the experimenters present when the data were collected (the present author). Details regarding the coding are provided in §4.6.3 and §4.6.4.

[8] This was following Umeda (2001), who used a series of pictures rather than the one large picture containing several different actions and objects which Whong-Barr and Schwartz used. This was more effective at eliciting data which was less fragmentary.

4.3 What to measure

In her book-length analysis of the linguistic proficiency of Turkish-speaking children acquiring Dutch, Lalleman (1986) argues that it is necessary to assess both (syntactic) complexity and (syntactic) accuracy as these two aspects of language proficiency closely interact. Put simply, the more complex an utterance is, the more opportunity there is for errors to be made. If an assessment of L2 proficiency were made on the basis of complexity alone, where complexity is measured, for example, by length, L2ers who produce long yet inaccurate utterances would be considered more proficient than L2ers who produced shorter but more accurate utterances. This holds of lexical as well as morphosyntactic proficiency: a balance should also be struck between the production of a wide range of lexical items and the use of such lexical items in a targetlike fashion. It is therefore essential that both complexity and accuracy are incorporated into the proficiency score.

Larsen-Freeman (1978a; 1978b; 1983) calls for an 'index of L2 development' comparable to MLU (mean length of utterance) used in L1 acquisition research. She writes that such a yardstick should:

> give a numerical value to different points along a second language development continuum – numerical values which would be correlates of the developmental process and would increase uniformly and linearly as learners proceed towards full acquisition of a target language (Larsen-Freeman 1978a:440).

In a series of studies on the written and oral production of adult ESL students, Larsen-Freeman evaluates a number of measures of complexity and accuracy. Most of these are based on the notion of T-unit, which is used to divide L2ers' output into countable parts. A T-unit is defined as 'one main clause plus whatever subordinate clause and nonclausal expressions are attached to or embedded within it' (Hunt 1970:14).[9] The measures which Larsen-Freeman examines include: average length of T-unit and total number of errorfree T-units (Larsen-Freeman and Strom 1977), percentage of errorfree T-units and average length of errorfree T-units (Larsen-Freeman 1978a; 1978b; 1983).

As Larsen-Freeman herself notes, measures such as the total number of errorfree T-units are task-dependent, that is, they vary according to how many utterances a subject produces on a given task. They are therefore relatively uninformative. The remaining three measures, average number of words per T-unit, average number of words per errorfree T-unit, and percentage of errorfree T-units were, to varying degrees, all found to be relatively successful in differentiating (holistically determined) proficiency levels. Percentage of errorfree T-units was found to consistently obtain the highest correlations with

---

[9] T-unit is the abbreviation of 'minimal terminable unit'. This term, based on research on written data (see below), is derived from the shortest grammatical unit (minimal) which can be punctuated as a sentence (terminable) (Hunt 1970:4).

proficiency level and discriminatory power, followed by average number of words per errorfree T-unit and average number of words per T-unit. Larsen-Freeman (1983) concludes that whereas mean length of T-unit (MLU) may be a satisfactory index for L1 acquirers, it seems that mean length of *errorfree* T-unit is more appropriate for L2 acquisition. 'By assessing a learner's *errorfree* T-units, a learner's score cannot be artificially inflated by his or her simply stringing words together' (Larsen-Freeman 1983:295, original emphasis). Larsen-Freeman thus decides that, consonant with Lalleman's observations, a combinatory measure based on complexity *and* accuracy is the most successful means of measuring L2 proficiency.

The proficiency score which is adopted here also combines measures of complexity and accuracy. For reasons which are outlined below (§4.4.1), MLU (either of all utterances or of errorfree utterances only) was found to be unsuitable for present purposes. The score which is used here is based on three sub-measures: (i) verbal density (the number of finite and non-finite verbs per T-unit) as a measure of morphosyntactic complexity, (ii) lexical richness (Guiraud's index ($V/\sqrt{N}$)) a measure of lexical complexity and (iii) rate of errorfree utterances as a measure of morphosyntactic/lexical accuracy. Section 4.3.2 provides an overview of how exactly these measures are operationalised and how they are combined to form a single proficiency score for each subject. Prior to this, in §4.3.1, I discuss the types of utterances which were excluded from analysis. Here, the discussion also deals with how the notion of utterance should be defined and how to deal with mixed utterances, that is, utterances which contain elements in both Dutch and English. The remainder of the chapter is dedicated to a detailed motivation of why the particular measures selected for inclusion here were chosen above others.

4.3.1 Types of utterance to include and types of utterance to exclude

This section deals with how the notion of utterance should be defined and which utterance types should be included when calculating measures of complexity and accuracy. Much of the research reviewed here deals with the use of MLU in L1 acquisition and most of the examples are based on this. The issues raised, however, are applicable to any measure which requires the subjects' samples to be divided into countable units.

The types of unit which have been used as a basis for calculating MLU and equivalent measures include the utterance, the clause, and the T-unit. Using different units of measurement as a basis for calculating complexity measures clearly undermines the cross-study comparability of any findings based on these calculations. A failure to clearly operationalise the notion of utterance was, amongst other things, one of the first criticisms of Brown's original MLU measure (Crystal 1974). Chapman defines utterance using 'the criterion of terminal intonation contour, rising or falling' (1981:22). In his study of narrative organisation in young children, Aarssen (1996:28) uses the following (rather vague) definition of clause: 'a clause is a unit in which information about events, activities or states is coded.' Berman and Slobin (1986:37; cited in Slobin 1993) define the clause as

> any unit that contains a *unified* predicate. By *unified*, we mean a predicate that expresses a *single* situation (activity, event, state). […] In general, [..] treat as a single clause those utterances that have two verbs but one subject, and treat as two separate clauses cases when each verb has a different subject – e.g. ***I** want to go* vs. ***I** want **you** to go.*

The example given in this latter definition would result in ungrammatical clauses (such as *I want* and *you to go*). This renders it unsatisfactory for present purposes because accuracy (grammaticality) will also be calculated as part of the proficiency measure. Measures of complexity and accuracy can also be calculated using T-units (see above for definition). For example, Bol and Kuiken (1988) base their MLU calculations for L1 Dutch on T-units. In the present study, for reasons which will become clear in §4.4.4, T-unit is also used as the relevant unit of measurement.

Most of the discussion on how to calculate MLU focuses on the type of utterances which should or should not be included in a sample. Brown (1973) calculated MLU on the basis of 100 utterances, taken from the second page of the transcription and excluding any partly unintelligible utterances and filler words such as *mm* and *oh*. The words *hi, yeah* and *no* were included, as were utterances which contained best guesses and exact repetitions. The number of morphemes was then divided by the number of utterances.[10] In his calculation of MLU, Chapman (1981) included and excluded the same type of utterances as Brown, but he used only 50 utterances and these were taken from the beginning of the transcript.

Chapman (1981) emphasises that in order to obtain a representative MLU value for a given child, it is important to ensure that the sample of data on which the MLU calculation for that child is based is representative. That is, it should exclude potential biases such as a high rate of imitations, frequent self-repetitions, a high rate of elliptical answers, frequent routines such as rhymes, etc. or a high rate of utterances in which clauses are conjoined with *and*. As Chapman (1981) goes on to note, using the T-unit as the basic unit of measurement circumvents this latter bias.

Johnston (2001) presents an alternative method of calculating MLU which avoids many of the potential biases Chapman refers to. In her calculation, she excludes the following types of utterance:

---

[10] Brown (1973: 54) gives other rules for calculating MLU but as these deal with what should or should not count as a morpheme and since the discussion of MLU in §4.4.1 focuses mainly on MLU when measured in words, they are not relevant.

(1) a. Exact self-repetitions
b. Exact repetitions of the adult partner
c. Single-word responses of the following two types:
i. Yes/No (and equivalents, e.g. uhuh, nope), acknowledgements of the adult's previous utterance. Adult: "I'm going to put it there." Child: "Yes."
ii. Yes/No (and equivalent) answers to questions
d. *Wh* question responses that provided only a queried constituent. Adult: "Where should I put the moose?" Child: "In there."

(Johnston 2001: 158-159)

Johnston analyses spontaneous data from 24 normally-developing monolingual English-speaking children, aged 2;5 to 4;7, and for each child's sample of 250-300 utterances she calculates both the traditional and her alternative (MLU2) MLU values. She finds that for approximately 25% of the children, the difference between MLU and MLU2 (both measured in morphemes) is less than 10%.[11] For a further 20%, this difference is between 26% and 49%. A stepwise regression analysis (on the group data) indicates that the variable which accounts for most of this difference is the rate of child utterances that directly follow questions. This suggests that in its traditional form, MLU to a certain extent reflects the discourse properties of the sample in question, such as the number of question-answer exchanges, for example. As Johnston (2001:161) notes, '[b]ecause MLU is generally interpreted as a developmental index of language proficiency, variation in MLU that is due primarily to pragmatic factors could compromise its utility'. This point holds for any complexity measure based on a sample which includes the types of utterances listed in (1) above.

This finding is of import for the present study because some subjects' data include such question-and-answer sequences (often attempts on the part of the experimenter to elicit data, sometimes because the subject was shy, for example). To ensure comparability across subjects it is necessary to check the extent to which subjects vary with respect to the occurrence of the four types of utterance which Johnston excluded, and in particular, with respect to the rate of subjects' utterances which follow questions. As Johnston notes, such comparability can be achieved by either controlling for reply rates or using her alternate MLU2 (see also Vermeer 1986) for similar discussion). The rate of elliptical answers in response to questions was calculated for all subjects. On average, 4.8% of the utterances in each subject's sample involved such utterances. This figure ranged from 0% to 22%. Given this relatively large range, it was decided that in order to make the samples equivalent for each

[11] This difference is measured by means of a proportional difference variable, which takes the magnitude of the original MLU into account: %DIF = (MLU2-MLU)/MLU.

subject, such elliptical utterances, produced in response to questions, should be excluded.[12]

Finally, on the issue of context, it should be noted that for any measure of grammatical complexity, more reliable results are obtained when a specific method of data collection is used because this removes some of the possible biases (Shriner 1969). This is one clear advantage of the proficiency measure used in this study. The sets of pictures used were the same for all subjects, and across the two experiments, they were of a comparable nature. Furthermore, one of the two experimenters was constant across both sets of experiments (the present author) and the other experimenter (a native-speaker student assistant) was always constant within each set of experiments.[13]

One additional issue which arises when complexity measures counting the number of (certain) elements per utterance are used in L2 acquisition research is what to do with mixed utterances, that is, utterances which contain items from both the L1 and the TL. Logically, there are three options: (a) exclude all such utterances, (b) calculate the complexity measure on the basis of TL items only, that is exclude all L1 items from this calculation, or (c) include such utterances. The last option is clearly the simplest in execution as it requires no extra coding. The risk with this option, however, is that a subject's final complexity measure value may be artificially inflated by the inclusion of non-TL items. However, simply excluding L1 items only (option (b)) is also unsatisfactory: if the L2er uses these L1 items because of missing lexical knowledge, excluding them from the complexity measure calculation could result in a lower value, which in turn would be assumed to be representative of a less developed/complex grammatical system. In this way, then, a lack of *lexical* knowledge would be mistakenly factored into the *morphosyntactic* complexity measure. Here, I am assuming that if subjects had known the relevant TL lexical item, they would have used it. In other words, the subject clearly intended for a word to be present in the utterance, and therefore, to use a calculation which implicitly assumes that this is not the case would be an inaccurate reflection of the L2er's interlanguage grammar. Furthermore, from a practical point of view, option (b), that is excluding all L1 items, would require considerable additional coding (by hand).

The ideal solution would therefore seem to be option (a): excluding mixed utterances. There is, however, one practical issue mediating against this,

---

[12] One possible (unfortunate) consequence of excluding these utterances is that the remaining number of utterances for one or two subjects becomes so low that the complexity and accuracy measures can no longer be sensibly calculated (see below and Appendix B for details). It is however quite likely that the number of utterances for these subjects would have been very low anyway.

[13] The task was also carried out with (adult and child) native-speaker controls and, as will become evident in the following sections, the data collected from these subjects will provide a crucial source of comparative data with which to evaluate the suitability of certain complexity measures.

namely that for some subjects, this would result in the exclusion of a large proportion – and for some, the majority – of utterances, with the consequence that the representativeness of the sample used to calculate grammatical complexity would be undermined.

Taking all things into consideration, option (c) is considered the optimal solution. The lexical richness measure, which counts these L1 items as non-targetlike, should hopefully counteract any risk of artificially inflating a subject's score by including them in the grammatical complexity calculation.

This rounds off the discussion of which types of utterances are excluded from analysis here. A complete list with examples is provided in §4.6.3.

4.3.2 The measure in a nutshell

The proficiency score used here is based on three measures, two measures of complexity and one of accuracy. 'Verbal density' is used as a measure of morphosyntactic complexity and 'lexical richness or diversity' is employed for lexical complexity. Morphosyntactic and lexical accuracy is calculated on the basis of the rate of errorfree T-units. An overview is given in Table 10. Each of these three measures is presented in turn.

Table 10. Overview of proficiency measure

<table>
<tr><th>Type of measure</th><th>Linguistic domain</th><th>Measure</th><th>Resulting sub-score</th><th>Final score</th></tr>
<tr><td rowspan="2">Complexity</td><td>Morphosyntax</td><td>Verbal density:

number of finite and non-finite verbs divided by total number of T-units</td><td rowspan="2">Scores converted into standardised (z) scores for L2 child and L2 adult groups separately</td><td rowspan="4">Sub-scores are combined into a single score using principal components analysis. Each subject has a single standard normal (z) score as final proficiency score</td></tr>
<tr><td>Lexicon</td><td>Guiraud's index:

$V/\sqrt{N}$
(where V=type and N=token)</td></tr>
<tr><td rowspan="2">Accuracy</td><td>Morphosyntax</td><td rowspan="2">Rate of errorfree utterances:

number of errorfree utterances divided by total number of T-units</td><td rowspan="2">%</td></tr>
<tr><td>Lexicon</td></tr>
</table>

4.3.2.1 Verbal density

'Verbal density' is used to denote a measure of grammatical complexity involving the average number of finite and non-finite verbs per T-unit. It is calculated by counting the number of verbs per T-unit and dividing this by the total number of T-units in a sample. For example, if a subject produces a total of 75 verbs across 50 T-units, the resulting verbal density score is 1.5. The verb count includes copula *zijn* 'to be', the auxiliaries *hebben* 'to have' and *zijn* 'to be', *gaan* 'to go' and modals, as well as lexical verbs. All verb forms, that is, finite, non-finite and stem forms are included.[14] More details are provided in §4.4.4.

The subjects' scores on this measure are as follows. For the proficiency task carried out as part of the production experiment (reported in Chapter 5), verbal density scores range from 1.00 to 1.65 verbs per T-unit for the L2 children (mean = 1.28; SD = 0.16) and from 0.94 to 1.96 verbs per T-unit for the L2 adults (mean = 1.43; SD = 0.31). For the proficiency task carried out as part of the comprehension experiment (reported in Chapter 6), verbal density scores range from 0.74 to 1.68 verbs per T-unit for the L2 children (mean = .30; SD = 0.21), and from 0.91 to 1.88 verbs per T-unit for the L2 adults (mean = 1.38; SD = 0.28). (Note that for the purposes of the proficiency task, all subjects who were younger than 14 at time of testing were classed as children and those older than 14 were counted as adults (see §4.4.1.2 for further discussion).)

The L2 data were compared with picture description data collected from L1 children (7-, 9- and 11-year-olds) and L1 adults using exactly the same task. This comparison suggested that the L2 children's failure to score as high as the L2 adults on this measure is the result of the age difference between these two groups rather than a difference in proficiency. Consequently, in order to ensure that the age factor does not adversely affect the proficiency score and how it is used to compare L2 children with L2 adults, the verbal density scores for the two (child and adult) L2 groups are converted into standardised (z) scores for each group separately, before being combined with the other sub-scores. The logic behind this calculation is detailed in §4.4.4.

4.3.2.2 Guiraud's index

Lexical complexity was determined using a variant of the Type/Token Ratio known as 'Guiraud's index' or the 'indice de richesse'. This is calculated by counting the number of different lexical types a subject produces (V) and dividing this by the square root of the total number of tokens produced by that subject (N) i.e. $V/\sqrt{N}$. Subjects who produce 350 tokens and 100 different types, for example, will be assigned a lexical diversity score of 5.35 ($100/\sqrt{350}$).

The subjects' scores on this measure are as follows. For the proficiency task carried out as part of the production experiment, the lexical diversity

---

[14] The fact that many of these are non-targetlike (either in form and/or position) is factored into the proficiency score via the accuracy measure.

scores range from 2.41 to 8.47 for the L2 children (mean = 5.25; SD = 1.43) and from 3.68 to 9.58 for the L2 adults (mean = 6.61; SD = 1.73). For the proficiency task carried out as part of the comprehension experiment, the lexical diversity scores range from 3.58 to 7.54 for the L2 children (mean = 5.34; SD = 1.08), and from 3.96 to 8.94 for the L2 adults (mean = 6.89; SD = 1.55). Once again, comparable data from L1 children and L1 adults who carried out the same task suggest that the L2 children's failure to score as high as the L2 adults is a result of the age difference between these two groups rather than a difference in proficiency. Therefore, as with the verbal density scores, the two groups' lexical diversity scores were converted to z-scores for each group separately before being combined into the overall proficiency score. This serves to minimise the age effect on the child L2 ~ adult L2 comparison.

4.3.2.3 Rate of errorfree utterances

The final component of the proficiency score is rate of errorfree utterances. This measure of accuracy is calculated by dividing the number of errorfree utterances a subject produced by the total number of T-units. A complete list of the morphological, syntactic and lexical errors which were counted, as well as those which were excluded, is provided in §4.6.

The subjects' scores on this measure were as follows. For the proficiency task carried out as part of the production experiment, the rate of errorfree utterances ranged from 8.3% to 89.3% for the L2 children (mean = 44.1%; SD = 23.4%), and from 16.7% to 78.0% for the L2 adults (mean = 54.9%; SD = 16.4%). For the proficiency task carried out as part of the comprehension experiment, the rate of errorfree utterances ranged from 12.5% to 96.5% for the L2 children (mean = 54.8%; SD = 22.2%), and from 40.0% to 95.3% for the L2 adults (mean = 65.6%; SD = 17.0%).[15]

4.3.2.4 Computing a single score

The three sub-scores, verbal density, lexical diversity and rate of errorfree utterances, were combined into a single, overall proficiency score by means of a statistical procedure called 'principal components analysis'. Principal components analysis is a way of reducing the number of variables in a data set and of detecting structure between these variables. It extracts the commonalities, that is, the overlap between variables, such that as much of the variance as possible in these variables is accounted for. More details are provided in §4.7. The result of the principal components analysis is a set of scores with a standard normal distribution. These standard normal (z) scores constitute the final, overall proficiency score for each subject. For the proficiency task carried out as part of the production experiment, they range

[15] There are no comparable L1 data for this sub-score as the native speakers performed at ceiling.

from –2.06 to 2.12 and for the proficiency task carried out as part of the comprehension task, from –1.95 to 1.70. As these are standard normal (z) scores, this means that 0 is the average score, positive scores are above average and negative scores are below average. On the basis of these scores, subjects will be divided into different proficiency level groups in Chapters 5 and 6.

4.3.2.5 Overview of remainder of chapter

The remainder of this chapter provides a detailed account of why the measures employed here were selected and why other, perhaps more obvious, alternatives were not. The discussion includes a detailed review of the available literature on the reliability and validity of MLU as a measure of complexity, especially in children older than five and in L2 adults (§4.4.1). It also outlines why Type/Token Ratio was considered an unsuitable measure of lexical diversity (§4.5), and finally, it provides a detailed list of which types of errors were and were not included in the accuracy measure (§4.6). For the reader who is perhaps less interested in such methodological issues, it is also possible to move directly to Chapter 5 at this juncture.

4.4 Measuring morphosyntactic complexity

Following Ortega (2003:492), (morpho-)syntactic complexity is defined as 'the range of forms that surface in language production and the degree of sophistication of such forms'. As hinted at in the summary of Larsen-Freeman's work, this is usually measured by determining the average length of utterance, clause, or T-unit, either in words or morphemes. Additional measures include the average number of (subordinate) clauses per T-unit, the rate of verbal utterances and the measure which is adopted here, verbal density, that is, the average number of finite and non-finite verbs per utterance. The pros and cons of each of the aforementioned measures are summarised in §4.4.1 through §4.4.4. By far the most common of these measures in L1 acquisition is MLU. In the context of the present study, the use of MLU raises some interesting and important issues, primarily concerning whether this measure can be extended to use with (i) older children and adults, and (ii) L2ers. I will present data which show that because MLU still develops in older children and is highly variable in (native-speaker) adults, it is unsuitable for use with L2ers, at least for the purposes of the present investigation. Using comparative L1 data as a guide, I illustrate why verbal density proves to the better option as the measure of syntactic complexity.

4.4.1 Mean Length of Utterance (MLU)

In L1 acquisition, MLU (Brown 1973; Nice 1925) is the generally accepted means of assessing a child's stage of (morphosyntactic) development, that is, the level of grammatical complexity in a child's developing grammar.[16] Its widespread use may, however, be put down to the ease with which it can be calculated and the lack of a suitable alternative as much as to its validity as a measure of linguistic proficiency. As the frequent discussions witnessed in the literature demonstrate, both its validity and reliability remain disputed. In this section, I consider some of these issues in the context of the child L2 ~ adult L2 comparison which is the focus of the present study. These include: (i) whether MLU is a valid measure of grammatical complexity (§4.4.1.1), (ii) the applicability of this measure beyond the earliest stages of development, that is, for older (L1/L2) children and for (native/L2) adults (§4.4.1.2) and (iii) whether MLU is a reliable measure of grammatical complexity (§4.4.1.3).

MLU can be calculated in morphemes (MLUm) or words (MLUw). Both capture the increase in length resulting from increasing grammatical complexity, but using morphemes is more sensitive to the acquisition of inflection, allowing more fine-grained distinctions, particularly in the early stages of linguistic development. Such a detailed analysis is generally inappropriate for L2ers because from the start, most of their utterances are multi-morphemic. Furthermore, given that MLUw has been found to correlate with MLUm (Arlman-Rupp, van Niekerk de Haan and van de Sandt-Koenderman 1976 for Dutch; and Hickey 1991 for Irish), a more detailed analysis might be considered superfluous. Additionally, older L1 children, who are of a comparable age to the L2 children under discussion here, also produce multi-morphemic utterances and hence most of these studies consider MLUw only. The discussion here includes both studies which use MLUm and studies which use MLUw.

4.4.1.1 Assessing the validity of MLU as a measure of grammatical complexity

This section considers the extent to which MLU really does measure (morphosyntactic) complexity. One of the ways in which studies have attempted to establish the validity of MLU as a measurement of grammatical complexity is by establishing whether MLU correlates with other measures. For example, Scarborough, Rescorla, Tager-Flusberg, Fowler and Sudhalter (1991) assess to what extent MLUm correlates with the *Index of Productive Syntax* (Scarborough 1990); Klee and Fitzgerald (1985) compare MLU with the

---

[16] MLU is also used in L1 acquisition research to make crosslinguistic comparisons, that is, to compare the developmental trajectories of children acquiring different L1s. The crosslinguistic application of MLU (and of the proficiency measure as a whole) is not directly relevant here, but as Larsen-Freeman points out, 'it would be ideal if we could find a measure applicable to the acquisition of all second languages and one which would not be affected by a learner's native language background' (1978a:140). It is hoped that the measure developed here might go some way to filling this gap (although it would clearly have to be applied slightly differently on each occasion as the phenomenon under investigation in any experiments will have to be excluded from the picture description data in the same way as the utterances containing scrambling have been here).

Language Assessment, Remediation, and Screening Procedure ((LARSP) Crystal, Fletcher and Garman 1976).

The alternative measure used in the Scarborough *et al.*. study, *IPSyn*, is a grammatical complexity task which measures the emergence of various syntactic constituents, such as NPs and VPs. Given that these authors employ such a measure of emergence to determine the validity of MLU, it is perhaps unsurprising that the correlations they observe between this measure and MLUm are at their strongest when MLUm is less than three. The production of full NPs or full VPs, which is what is measured in the *IPSyn* score, will automatically lead to an increase in MLUm. The authors claim that the weaker correlation between MLUm and *IPSyn* score for MLUm > 3.0 indicates that MLUm values above this value are less closely associated with grammatical complexity and hence should not be used as a reliable estimate of a child's stage of linguistic development, nor should they be used to match individuals. It is unlikely, however, that such a conclusion is warranted, given that *IPSyn* measures the emergence of grammatical complexity only.

The alternative measure which is compared to MLU in Klee and Fitzgerald's (1985) study, *LARSP*, uses the relative frequency distribution of various clauses and phrases to assign children to different developmental levels. Klee and Fitzgerald analysed data from English-speaking children (n=18) between the ages of 2;1 – 3;11, who were divided into three groups of six (equivalent to Brown's stages) on the basis of their MLUm, which ranged from 2.5 to 3.85. The authors found that, on the whole, these MLU-based groups were very similar in terms of the developmental levels to which the children in each group had been assigned on the basis of *LARSP*. They conclude, therefore, that MLUm cannot predict diversity in syntactic constructions beyond a value of MLUm ≥ 2.5. Note, however, that for this analysis Klee and Fitzgerald use group data only, that is, they use the mean relative frequency distribution of each clause/phrase for the group. It is possible that these group means disguise important individual differences which may contribute to the lack of differentiation between the MLUm-based groups.

Another study seeking to establish the validity of MLU by determining whether it is a predictor of linguistic development as calculated using other measures is that of Rondal, Giotto, Bredart and Bachelet (1987). Rondal *et al.* found that MLUm *could* differentiate between the children in their study in terms of the relative frequency distribution and diversity of various grammatical constructions, *contra* Klee and Fitzgerald (1985). They conclude that MLUm can predict syntactic complexity up to an MLUm value of 3.0, which was the highest MLUm value for the children in their study. Blake, Quartaro and Onorati (1993) also find that MLUm can predict syntactic complexity, up to an even higher MLU value of 4.5. They examined data from 87 English-speaking children aged between 1;6 and 4;9. Using a similar method to that of Klee and Fitzgerald (1985) to measure syntactic complexity, they find a significant correlation between this measure and MLUm.

In an even earlier study, Shriner and Sherman (1967) report that MLUw (in their terms, Mean Length of Response) is the best single predictor of linguistic development. They examined 300 language samples taken from 200 children, aged 2;6 to 12 years (mean = 6;7), consisting of responses to picture stimuli and/or questions and remarks from interlocutors. Each sample was assessed with respect to a language development scale, ranging from "1" for least developed and "7" for most developed, by at least two out of 104 judges. Shriner and Sherman subsequently assessed the validity of MLUw (amongst other measures of grammatical complexity) by determining how well it correlated to the point on the language development scale to which the children had been assigned. Although a strong significant correlation was observed, the paper provides no details regarding the criteria judges used in order to assign children to the language development scale. It is highly likely that MLUw actually featured in the judges' decisions and, hence, the correlation between MLUw and this particular measure of linguistic development cannot be considered valid. This highlights a problematic aspect of any study which seeks to assess MLU on the basis of another language proficiency/development measure: the validity of this comparison ultimately depends on the validity of the alternative measure which is employed. For example, the comparison between MLU and the *IPSyn* measure carried out by Scarborough *et al.* can only be considered informative if one assumes that emergence is a valid criterion for acquisition.

Bearing this caveat in mind, the studies reviewed in this section suggest that MLU is a valid indicator of grammatical/syntactic complexity in the earliest stages of linguistic development, that is, up to a value of MLU 3 and possibly even later. The issue of whether MLU can be used after this point is of direct concern here, given that the subjects in the present study are much older than the children tested in the aforementioned studies.

4.4.1.2 Extending MLU beyond the earliest stages of development

In this section, I ask (i) whether there is evidence to suggest that MLU is a valid measure of L2 development and (ii) if there is, whether it can be regarded as a suitable means of comparison between L2 children and L2 adults. I will conclude that this is not the case.

4.4.1.2.1 MLU in L2 acquisition

L2ers who are older than L1 children are by definition cognitively more mature than the L1-acquiring child, quite simply because they are older. They also by definition know one other language already. Consequently, they are usually capable of producing multi-word/morpheme sentences almost immediately after initial contact with the target language (Adamson 1988; Larsen-Freeman and Strom 1977:124). This means that their initial MLU is comparatively high and hence there is less room for the L2er to develop in this respect. In other words, in terms of MLU, L2ers are likely to 'skip a stage' (or several) in comparison with L1 children. In this section, I consider whether there is evidence to show that MLU is nevertheless a valid measure of linguistic development in L2 acquisition. In a comparable fashion to the L1 studies reviewed in the previous section, most of the studies which address this issue look for correlations between MLU and proficiency level as determined by some other means.

Larsen-Freeman and Strom (1977) find a steady increase in the average length of T-unit (words) for written data from adult L2ers of English, although there were no significant differences for this variable between different (previously, holistically assigned) proficiency groups. As Dewaele (2000) points out, however, there lies a potential problem of circularity in Larsen-Freeman and Strom's (amongst others') line of argumentation: if the proficiency levels are based on holistic judgements made by assessors (such as teachers or the researchers themselves, as was the case in Larsen-Freeman and Strom's study), these judgements could subconsciously involve utterance length as a criterion for determining proficiency. It would therefore be unsurprising that, in such cases, a correlation should be found between MLU (per T-unit) and proficiency level (see also the discussion of Shriner and Sherman (1987) above).

Dewaele (2000) argues, on the basis of previous research, that MLU should not be used in L2 acquisition research, for the following reasons. Firstly, citing Lightbown (1977), he writes that MLU in L2 acquisition is subject to considerable variability: Lightbown found that MLU was not systematic in the sense that the most advanced L2ers in her study did not produce the longest utterances. It is worth noting, however, that Lightbown's subjects were child L2ers and hence it is possible that this lack of systematicity reflected differences which may be part of child (L1 or L2) linguistic development. Secondly, Dewaele notes that Blum-Kulka and Olshtain (1986) found that the L2 development of MLU is not linear. Specifically, these researchers observed that in making requests, low proficiency L2ers of English tended to use shorter utterances when compared to native speakers, whereas more proficient L2ers used *longer* utterances than the native speakers. Closer examination of this study reveals that it is however not clear (i) how the notion utterance was defined, and (ii) how proficiency level was determined. The import of these findings is therefore difficult to assess.

Dewaele (2000) goes on to argue that although MLU may be an inappropriate means of measuring diachronic variation in L2 acquisition, it can be used to measure synchronic variation. Dewaele collected three types of data from 21 Dutch-speaking adult L2ers of French: formal oral, informal oral and written. Rather unsurprisingly, he observes that MLU is longest in the written data and that it is longer in the formal oral data than in the informal oral data. An analysis of the distribution of the different lengths of utterances reveals a predominance of shorter utterances in the informal oral data. In the written data, the higher MLU value is not necessarily due to longer utterances per se (the same range of MLU values is found across all sets of data); rather, it is due to there being fewer shorter utterances. Dewaele goes on to propose an alternative measure, namely the MLU of the longest three utterances in a sample. He claims that this constitutes a better measure of the variability of L2ers to produce very long utterances. There are two problems with this alternative measure, however. Firstly, the choice of three (as opposed to four, five, etc.) utterances remains unmotivated. Secondly, MLU is supposed to measure the *mean* length of utterance, not the ability to construct long sentences. Thus, although Dewaele raises some valid issues regarding regular MLU, his alternative proposal does not fair any better, at least for the purposes of the present study.

Ortega (2003) synthesises 25 studies which examine the relationship between proficiency level and various measures of syntactic complexity, including mean length of T-unit. Using stem-and-leaf plots (Greenhouse and Iyengar 1994, cited in Ortega 2003), she assesses how large the mean difference between two proficiency groups (determined by programme level or holistic rating) must be in order for that difference to be statistically significant. For mean length of T-unit, this translates into the following question: How many words does the mean difference between two proficiency groups have to be in order for those two groups to be considered (significantly) different?

Ortega compares 68 cross-group comparisons from 19 studies, all of which examine written data from college-level L2ers. She finds that mean differences of two or more words tend to be significant, and if the sample sizes involved are large (20 to 50 participants per cell), this figure is slightly lower (1.8 words). Conversely, if the sample sizes are small (ca. ten participants per cell), this figure is slightly higher (2.2-2.9 words).[17] Mean differences of less than one word are less likely to be significant. Irrespective of the specific differences between groups, what is crucial in Ortega's findings is that they indicate that L2ers develop with respect to MLUw, although without examining each individual study included in Ortega's survey, we do not know

---

[17] Note that these cross-group comparisons are not always of the same type: most (49) involve two adjacent proficiency levels, 16 include maximally different proficiency levels (i.e. highest and lowest) and three were between an advanced group of L2ers and native speakers. For adjacent level comparisons, the mean difference was 1.4 words, and for maximal difference comparisons, it was three words. This disparity between mean differences suggests that the statistically significant cross-group comparisons, which all had mean differences of 2.2 words or more, were probably maximal difference comparisons rather than adjacent level comparisons.

at what level the L2ers' MLUw values actually are.[18] However, given that the (significant) differences between groups range from 2.2 words to 5.7 words, we can safely assume that the most proficient will be beyond at least MLUw = 7.0 words. Dewaele's findings reported above suggest that this value is likely to be slightly lower for the informal, oral data which are used in the present study.

Verhoeven and Vermeer (1989) evaluate how MLUw compares with standardised tests designed to test syntactic proficiency. They administered the *TAK Toets Allochtone Kinderen* 'Immigrant children test' to 20 7-year-old Turkish-speaking child L2ers of Dutch, with an average length of residency of 5.75 years.[19] In this test, syntactic proficiency is measured using two tasks. In a sentence comprehension task, subjects are asked to choose the picture from a series of three which best fits the sentence they have just heard. Example sentences are given in (2)-a and (2)-b.

(2)	a.	Welk meisje	heeft	geen	hoedje	op?	
		which girl	has	no	hat	on	
		'Which girl isn't wearing a hat?'					
	b.	Het	is	de jongen	die	het meisje	duwt
		it	is	the boy	that	the girl	pushes
		'It's the boy who is pushing the girl.'					

The task measures knowledge of function words and relations within and between words. In the sentence formation task, subjects are asked to imitate a series of sentences and they are scored on their correct realisation of function words and word order patterns.[20] Pronunciation errors and mistakes such as failing to realise the adjectival inflection are not counted. Example sentences, where the part of the sentence to be evaluated is given in bold, are given in (3). The sentence in (3)-b, for example, measures whether subjects know the target verb-final word order in subordinate clauses.

(3) a.	**Wie**	maakt	**de tekening**	**van Mustafa**	vies?
	who	makes	the picture	of Mustafa	dirty
	'Who is making Mustafa's picture dirty?'				

[18] Across 21 of the studies on L2 college writing which Ortega discusses, the MLU (per T-unit) was 12.18 words (SD 3.03; Range 5.10 – 18.40). Note that the written mode undoubtedly contributes to this high score (cf. native-speaker scores in the next section).

[19] It should be noted that the purpose of the Verhoeven and Vermeer (1989) study is quite different from the purpose for which their findings are used here. They compare *TAK* scores and equivalent spontaneous data scores with teacher evaluations and they use the latter evaluations as a basis for evaluating the *TAK* and spontaneous speech scores.

[20] This is the task used by Blom and Polišenskà (to appear) (cf. fn. 7).

b. **Ik** **moet** **dan** stoppen bij het eerste huis
I must then stop at the first house

dat **na** **de** **hoek** **komt**
that after the corner comes
'I have to stop then at the first house after the corner.'

The assumption behind such imitation tasks is that when a particular aspect of language is not part of a given subject's developing grammar, s/he will not be able to repeat it (Bley-Vroman and Chaudron 1994; Lust et al. 1996).

From the same children, Verhoeven and Vermeer obtained a corpus of spontaneous data (200 utterances for all children except two) using a picture description task and by asking questions on the topics of friends, television and holidays. Utterances consisting of 'yes/no' only, idiom chunks, interjections and direct imitations of the experimenter were excluded. Subsequently, using similar thinking to that of Dewaele (2000), Verhoeven and Vermeer calculated the MLUw for the longest 10% of utterances, which they consider to provide an indication of a subject's maximum verbal capacity. They find that this MLUw correlates significantly with both of the relevant *TAK* sub-tests: for the sentence comprehension task, $r = .56$, $p < .01$ and for the sentence formation task, $r = .59$, $p < .01$. This suggests that MLUw (of the longest 10% of utterances) measures something which is comparable to the syntactic components of the *TAK* test, and on the assumption that these parts of the *TAK* test are valid measures of syntactic proficiency, it may be concluded that MLUw (of the longest 10% utterances) is also valid in this respect.

Another adaption of MLU in L2 acquisition research is the mean length of *errorfree* T-unit. As noted in §4.3, Larsen-Freeman (1983) concluded that this was the most satisfactory measure of grammatical complexity in L2 development. This adaptation of MLU is less than ideal, however, because as Larsen-Freeman and Strom (1977) point out, there may be populations – for example, very low proficiency subjects, who were not tested in any of Larsen-Freeman's studies – for whom the number of errorfree T-units is so low that any assessment of proficiency based on this calculation might not be truly representative of these subjects' L2 competence because a substantial proportion of their total output would have to be discarded. This contributed to Whong-Barr and Schwartz's (2002) decision not to use this measure either.

To summarise: The few available studies on MLU in L2 acquisition give mixed conclusions with respect to the use of this measure as an indicator of grammatical complexity. On the one hand, it has been argued that MLU is a valid measure in L2 acquisition because it develops linearly with increasing proficiency level and it correlates with standardised tests. On the other hand, however, it has been claimed that such results may stem from circular argumentation where MLU is (subconsciously) included in how the different proficiency levels are determined.

4.4.1.2.2 MLU in older L1 children

This section examines the extent to which MLU develops in older L1 children. Establishing whether this is the case will allow us to determine whether MLU is a suitable measure for the child L2 ~ adult L2 comparison. The logic is as follows. Given that adults are older than children, it is to be expected that these two groups will probably differ in terms of their MLU. In and of itself, this difference need not be a problem (cf. the verbal density and lexical diversity measures). What would be problematic, however, in the sense that it would complicate the child L2 ~ adult L2 comparison, is if age were found to correlate significantly with MLU *throughout* childhood, that is, beyond the age of five, the age until which it has been established that such an age~MLU correlation exists (Johnston 2001; Miller and Chapman 1981; Rondal *et al.* 1987). The reason this would be problematical is because, whereas for L2 adults, MLU would solely be a measure of language development, for L2 children, it could also be a function of age itself, and this would introduce an unwanted additional variable to the child L2 ~ adult L2 comparison. Implicit in this argument is the assumption that the L2 children's MLU in their L2 will not progress beyond the value of their MLU in their L1 (which, I believe, is relatively uncontroversial when exposure to the L1 is maintained, as in the present study).

The following example serves to illustrate this argument. Imagine we have two L2 children, Dom and Helen. Both are native speakers of English. Dom was first exposed to Dutch at age four and Helen at age seven. They are tested three years later, that is, when Dom is seven years old and Helen is ten. Assuming that all other variables are held constant, imagine that Helen's Dutch MLU is observed to be higher than Dom's. If MLU is still developing in L1 children between the ages of five and ten years, for example, it would be unfair to conclude that Helen's interlanguage is more complex than Dom's, because this might be due to her relative maturity. In other words, the reason for Dom's lower MLU in his L2 Dutch might not be because he is less proficient than Helen; rather, it might result from him being three years younger than she is and from his L1 MLU being lower, which, in turn, would mean that his L2 MLU would be lower. Before considering how such a confound between L1 MLU and L2 MLU might be dealt with, it must first be established whether there is evidence to suggest that MLU develops in L1 children beyond five years old. In this section, I review the available literature before examining this question using child L1 data specifically collected for the purpose.

Shriner (1967) reports that MLUw cannot be used as an indicator of linguistic development beyond the age of approximately five because of an increase in variability in MLU scores after this age. This contrasts, however, with Shriner and Sherman (1967), who found that in their data from 200 children between the ages of 2.5 and 12 (mean 6;7), MLU was the best single predictor of linguistic development. No information is given about the distribution of the different age groups or the specific MLUw values, however.

Chabon, Kent-Udolf and Egolf (1982) collected 100 utterances per day over a three-day period from three different groups of English-speaking children, all of whom were beyond Brown's Stage V: 3;6-4;6 (n=10), 5;6;-6;6 (n=10) and 8;6-9;6 (n=10). Although their analysis showed a main effect of age ($F = 3.77$; *df* = 2.24; $p < .05$), a Scheffé post-hoc comparison did not reveal any pairwise significant differences, that is, no single pair of age groups differed significantly from each other. The mean MLUm values across all three days were 5.95 for the 3-4-year-olds, 6.66 for the 5-6-year-olds and 6.81 for the 8-9-year-olds. As we shall see below (§4.4.1.3), considerable variation was observed within-subjects across different days. Minifie, Darley and Sherman (1963) also document the MLUw in 5-year-old (n=48) and 8-year-old (n=48) English-speaking children. Averaging MLUw values across three different sample moments, they observe a clear difference between the two groups: average MLUw for the 5-year-olds is 5.96 and for the 8-year-olds 8.96. They do not report whether this difference is statistically significant.

In a large-scale longitudinal study on various aspects of linguistic development in English-speaking children and adolescents aged six through 17, Loban (1976) observes a clear increase in MLUw (per T-unit) with increasing age. Oral data were collected using interviews on a variety of topics from the same 211 participants for 12 years. The results (for a randomly selected subset of 35 subjects) are presented in Figure 1. Figure 1 shows a steady increase in MLUw with increasing age, but without any statistical analysis, which Loban fails to provide, it is not possible to determine which age groups differ from each other.

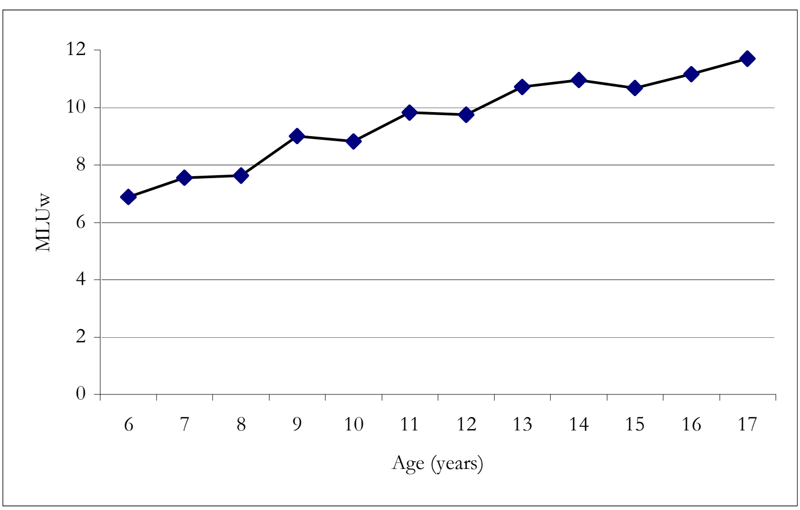


Figure 1. Loban (1976): MLUw (per T-unit) as a function of age

Hunt (1970) also observes a steady increase in MLUw (per T-unit) up to age 12. He analysed written data from 250 children and adolescents aged nine through 17, observing a significant increase with age (ANOVA, $p < .01$). Interestingly, he also finds that variability also increased with age. Compare the standard deviations for MLUw for the 9-, 11-, 13-, 15- and 17-year-olds, respectively: 1.13, 1.54, 3.06, 2.11 and 2.54. As suggested above, the extent to which these written data are comparable to spoken data is questionable. The nature of the task produced data which cannot really be considered to be parallel to spontaneous data: subjects were asked to rewrite a passage which had been written in extremely simple sentences. In other words, they were asked to make short sentences larger. The expectation, which seems to be borne out in the data, was that the more cognitively mature (older) subjects would combine more frequently, with the consequence that they would produce longer T-units. Such writing skills are also often taught at school.

The results of the studies presented in this section are mixed, although in general they suggest that MLU continues to develop after age five. Minifie *et al.* (1963), Loban (1976) and Hunt (1970) observe continuing development in MLU beyond age 5, whereas Chabon *et al.* (1982) find no significant differences between the age groups they examine. It is also not clear whether MLU values vary more with increasing age: Shriner (1967) and Hunt (1970) report that this is the case, whereas Shriner and Sherman (1967) suggest the contrary.

The mixed nature of these results make it difficult to draw any firm conclusions regarding the nature of MLU in older L1 children and its potential implications for the L2 children in the present study. Fortunately, however, data from the present study are available from child L1 control groups who carried out exactly the same task as the L2 subjects. In the remainder of this section, I analyse these child L1 data, comparing the different age groups with each other and with adult L1 data.

As part of the experimental procedure reported on in Chapter 6, data were collected from 72 L1 children aged between 7- and 12-years old. These picture description task data were transcribed by two trained native-speaker student assistants for ten randomly selected children in the 7-, 9- and 11-year-old age groups and for ten native adult controls. The adult data are included for comparison but they will be discussed in more detail in the following section, on reliability of MLU. Each student assistant transcribed half of the children and checked the transcriptions of the other half against the recordings. In order to make the data completely comparable with the L2 data, the types of utterance listed in §4.6.3 were excluded before MLUw was calculated for each subject. Recall that the logic of analysing these data is to determine whether MLU develops in L1 children beyond the age of five in order to ascertain whether there may be a potential confound between L1

MLU and L2 MLU for the child L2 subjects. The results are presented in Table 11.

Table 11. L1 child and adult MLU

Age group	N	Mean	Range	SD
7-year-olds	10	5.88	4.88-6.74	0.67
9-year-olds	10	6.41	5.49-7.39	0.60
11-year-olds	10	6.60	6.19-7.39	0.36
Adults	10	7.50	5.86-10.02	1.32

As expected, there is a significant difference between groups when the adults are included (Kruskal-Wallis (data converted to ordinal scale): $\chi^2 = 12.96$, df = 3, $p < .01$).[21] There is also a significant difference when the adults are excluded (Kruskal-Wallis (data converted to ordinal scale): $\chi^2 = 6.068$, df = 2, $p < .05$), which suggests that MLUw does develop significantly between the ages of seven and 11, that is, the different age groups within the L1 children cannot be considered as one group with respect to MLUw. The implication of this finding is that, on the logic laid out above, and assuming that L1 English-speaking children will pattern similarly to L1 Dutch-speaking children, using MLU as a measure of morphosyntactic complexity would introduce an additional variable into the child L2 ~ adult L2 comparison: for L2 adults, it would measure language development, whereas for L2 children it would measure language development and it would be a function of age. There are two ways to deal with this complication. First, each L2 child could be compared with an L1 child of the same age. One problem with this option, however, is that comparable L1 data are not available for all the relevant age groups. Second, it might be possible to somehow factor out the effect of the L1 MLU on the L2 children's L2 MLU. The problem with this option is that these data are not available for the L2 children and although it would in principle have been an option to test the L2 children in their L1, too, this was not possible for practical reasons.

Imagine, however, for argument's sake, that it were possible to factor out the L2 children's L1 MLU. In order for such a calculation to make sense, it would first need to be established that MLU were a reliable indicator of grammatical complexity in the first place. This question is addressed in the following section.

[21] It was not possible to carry out an ANOVA with these data because they did not conform to the assumption that the variances within each group were equal, as shown by Levene's test of homogeneity of variances: Levene's statistic = 5.136, df1 = 3, df2 = 36, $p < .01$.

4.4.1.3 Assessing the reliability of MLU as a measure of grammatical complexity

Assessing the reliability of MLU requires determining the extent to which it varies within and between both groups and individuals. If MLU is too variable, it must be considered unsuitable as a measure of grammatical complexity.

4.4.1.3.1 Intra-speaker variability

As well as investigating the validity of MLU as a measure of linguistic development, Klee and Fitzgerald (1985) also examine its reliability. They do this by calculating the standard error and 98% confidence interval of the MLUm for the transcript of every child in their sample. This involves drawing each possible block of the same number of utterances from each transcript and calculating the MLUm for each of these blocks, resulting in a set of MLUm values. The mean of these MLUm values provides an estimate of the true population mean, and the standard deviation of these MLUm values, i.e. the standard error, estimates the extent to which the sample MLUm differs from the true, unknown MLUm. Subsequently, it is possible to use the standard error to calculate the range of values in which we can be 98% confident that the true value will fall, that is, the so-called 98% confidence interval. For example, for a child whose 98% confidence interval is 3.45 – 4.22, we can expect that in 98% of the blocks we have taken from the language sample in question, the MLUm will fall somewhere between the values of 3.45 and 4.22. Klee and Fitzgerald find that for each child, the observed confidence interval spans two or three of Brown's stages (for five and 13 children, respectively). They conclude, therefore, that the observation that a child's MLUm value may vary considerably as a function of the 100-utterance block it is taken from undermines the value of MLUm as a means of estimating that child's stage of linguistic development.

Rondal *et al.* (1987) also examine the reliability of MLUm within children. They calculate confidence intervals for each individual child using exactly the same method as Klee and Fitzgerald. All but one of the 21 children in their study span two stages; in other words, there is less intra-subject variability in their data than in Klee and Fitzgerald's. The confidence interval does increase with age, however, as does the overlap between stages. Rondal *et al.* conclude that beyond an MLUm value of 3.50, variability becomes larger, and hence at these later stages MLUm may be a less successful measure of grammatical complexity. They do warn, however, that the individual distributions on which these confidence intervals are based are not normally distributed and these latter calculations should therefore be treated with caution.

Chabon, Kent-Udolf and Egolf (1982) examine the extent to which MLUw is reliable over time. Using the same sets of stimulus materials, they collected a sample of 100 utterances per day over a three-day period from children in three different age groups (3;6-4;6, 5;6-6;6 and 8;0-9;6). There were only low correlations between the MLUw values across the three days: between days one and two, r = .416, between days two and three, r = .330 and between days one and three, r = .330. All correlations were significant at the .05 level, but not at the .01 level. The correlation within groups was also low (r = .248), although when the average across all three days was used, the correlation became stronger (r = .577; see Minifie, Darley and Sherman (1963) for similar findings). Note that the day-to-day reliability for the two younger groups was much better than for the eldest group: 3-4 year olds, r = .422; 5-6 year olds, r = .407; cf. 8-9 year olds, r = .053. The authors conclude that any evaluation of children's linguistic ability based on one sample only should be treated with caution given that it is possible that the value obtained may reflect temporal instability as well as linguistic development (or lack thereof). Although this warning seems well-founded, it could equally be applied to any measure which makes use of spontaneous data.

Minifie *et al.* (1963) use a similar methodology to that of Chabon *et al.* (1982) to evaluate the temporal reliability of MLU. They also find that this measure is unstable over time: the 5-year-olds (n = 48) scored an average of MLUw 6.79, 5.70 and 5.39 on days one, two and three, respectively, and the 8-year-olds (n = 48) had an average MLUw of 10.31 on day one and 7.60 on day two. It is possible, however, that these discrepancies arose from the use of different stimulus materials on the different days. Hence, these findings should be treated with caution.

The studies reviewed here indicate that as a measure of grammatical complexity, MLU is quite variable, particularly beyond the earliest stages of development. MLU values may also vary within individuals as a function of testing occasion and sample.

4.4.2.2.2 Inter-speaker variability

In the remainder of this section, our attention turns to the issue of inter-speaker variability. For this, we consider MLU in native-speaker adults. Native-speaker adults are the target to which we compare (adult) L2ers and if it observed that as a group they are highly variable in this domain, this would undermine the usefulness of MLU as an indicator of linguistic development. Clearly, there will be *some* variation, but the question is how much. If native speakers are all observed to have an MLU above a certain level and the range of scores is relatively small, then we may use this as a measure of targetlike ability. If, on the other hand, there is considerable variation in their MLU values, MLU cannot be considered to be a reliable measure of grammatical complexity, at least not by itself.

The native-speaker data given above in Table 11 were used to assess the extent to which MLU varies across adult native speakers. To recap, the MLUw values for these ten speakers ranged from 5.86 to 10.02. The mean was 7.50 and the standard deviation was 1.32. Most (80%) of the native-speaker adults had MLUw values above 6.5, indicating a reasonable amount of consistency across speakers. Nevertheless, the large range of five words indicates that even when native-speaker adults perform the same task, their MLUw may vary considerably. This finding has two implications. Firstly, it suggests that MLUw as a measure of grammatical complexity for native-speaker adults is not very reliable. Secondly, as a consequence, pinpointing a value at which an L2er's MLUw value can be considered native-like would be rather difficult.

4.4.1.4 Summary on MLU

Although previous research has shown that MLU is a reasonably valid indicator of grammatical complexity in early child L1 acquisition and to a certain extent in L2 acquisition as well, this measure remains, on the whole, rather unreliable (across different measurement points and samples). An analysis of new L1 data from older children suggests that MLUw may develop well into the first decade of life. This is a potentially interesting finding in and of itself, although further research (for example with more subjects) is of course necessary to confirm it. In the context of the present study, however, where L2 children are compared with L2 adults, this result, coupled with the observation that even on the same task, MLUw in adult native speakers is highly variable, means that MLUw must be regarded as an unsuitable measure of grammatical complexity.

The remainder of this section outlines possible alternative measures: the rate of subordination, the rate of verbal utterances, and finally, verbal density, which, as indicated in §4.3.2, is the measure of morphosyntactic complexity adopted here.

4.4.2 Rate of subordination

The extent of subordinate (or dependent) clauses produced by a subject has also been used as a measure of grammatical complexity. This has been calculated in two ways: either by dividing the number of subordinate clauses by the total number of utterances (to obtain the rate of subordination) or by dividing the total number of (main and subordinate) clauses by the total number of utterances (to obtain the average number of subordinate clauses per T-unit).[22]

Hunt (1970) used the latter measure (average number of subordinate clauses per T-unit) to analyse the written data he collected from L1 English-speaking subjects aged nine through 17 (cf. §4.4.1.2.2). He observed a clear increase in subordination as a function of age: the younger subjects (9- and 11-year-olds) had an average of less than 1.2 subordinate clauses per T-unit, whereas the older subjects' (13-, 15- and 17-year-olds) scores were considerably higher at around 1.4. As noted above, however, the task used to obtain these data more or less required subjects to combine shorter sentences. It is therefore not clear how representative these data are of 'normal' rates of subordination in written data, let alone in spoken data. In his discussion of this and other measurements of subordination, Loban (1976) notes that only subordinate clauses containing finite verbs are counted in Hunt's calculation, and as a consequence, other grammatically complex forms, such as the gerund and non-finite dependent clauses, are missed. This is one reason why, in the present study, verbal density is adopted as a measure of morphosyntactic complexity.

Verhoeven and Vermeer (1989) used the rate of subordination in their comparison of spontaneous data and the standardised *TAK* test as a means of assessing L2 proficiency. In the data of their 20 Turkish children, they find that this particular measure did not correlate with either of the two syntactic measures in the *TAK* test (Pearson's $r = .19$ for the sentence comprehension task and .16 for the sentence formation task). It did, perhaps unsurprisingly, correlate with MLUw (of the 10% longest utterances; $r = .51$, $p < .01$), however.

These two studies suggest that rate of subordination and average number of subordinate clauses per T-unit are probably unsuitable measures of morphosyntactic complexity. An analysis of the child L1 and adult L1 data collected as part of the present study (analysed above in the discussion of MLU) provides further support for such a conclusion. A considerable proportion of the children produce very few subordinate clauses at all: no fewer than ten of the 30 children produce no subordinate clauses whatsoever and fewer than half of all subjects (19/40) have a score which exceeds 0.05

---

[22] The use of either of these measures clearly requires the utterance to be defined in such a way that dependent clauses are counted as belonging to the same utterance as the main clause on which they depend, e.g. as a T-unit.

subordinate clauses per T-unit. (The maximum score was 0.16 for the L1 children and 0.27 for the L1 adults). This means that the target native-like level for the L2 subjects would be very low, with the consequence that there would be very little room to measure development, which is one of the goals of the proficiency measure as a whole. For these reasons, the rate of subordination must be considered unsuitable as a measure of grammatical complexity in the present study.

4.4.3 Rate of verbal utterances

Another measure of linguistic complexity used in the L1 literature is the rate of verbal utterances, that is, the number of utterances containing verbs divided by the total number of utterances. Valian (1991) proposed this measure as a means of overcoming the problems encountered with MLU when comparing children acquiring a morphologically rich language with children acquiring a morphologically poor language (see Belletti and Hamann 2000 for its application to bilingual L1 data; see also for example Deen 2002 for L1 Swahili). Valian observes a clear increase from 10% to 84% in the rate of verbal utterances in children acquiring English between the ages of 1;10 and 2;8 (MLU 1.53 and 4.38). Unfortunately, such a measure is unsuitable for L2ers because their knowledge of another language means that they generally do not pass through the one-word/two-word/multi-word stages typical of L1 acquisition (Adamson 1988; Larsen-Freeman and Strom 1977:124) and, consequently, the stages for which this measure would be most useful are generally skipped over. This is confirmed by the observation that the L1 children in Valian's study are already at ceiling level at around age 2;8. As a result, it must be concluded that this measure is also inappropriate for present purposes.

4.4.4 Verbal density

In the final part of this section, I provide the background and rationale to the measure of morphosyntactic complexity which is adopted here, namely, verbal density, which is defined as the average number of finite and non-finite verbs per T-unit. This measure was adopted by Chaudron and Parker (1990), following Pica and Long (1986), in an analysis of free and elicited production data in Japanese-English adult L2 subjects. Pica and Long (1986) use verbal density in their analysis of conversations between native speakers and between native speakers and non-native speakers, but neither they nor Chaudron and Parker (1990) motivate their choice of this measure with evidence that it is a valid measure of L2 proficiency (morphosyntactic complexity). Note, however, that such a lack of construct validity is a problem with most of these measures, particularly in L2 acquisition research: noone (to my knowledge) has systematically investigated whether and how such measures of morphosyntactic complexity relate to developmental sequences. This means that any such measure faces this criticism and as such, it cannot be used as an argument against one over the other (Ortega 2000).

Despite the above-named authors' failure to motivate verbal density as a suitable measure, there are several reasons to believe that it is a valid means of measuring grammatical complexity, at least in a rudimentary fashion. First of all and most importantly, it captures complexity in a central aspect of grammatical development, that is the use of different verb forms, such as non-finite dependent clauses ((4)-a), relative clause modification ((4)-b), modals and complex constructions and tense forms ((4)-c and (4)-d), and durative constructions ((4)-e and (4)-f).

(4) a.

Ze	beslist	om	te	gaan	zwemmen
she	decides	for	to	go	swim

'She decides to go and swim.'

b.

Het meisje	dat	een groene trui	aan heeft	valt	om
the girl	that	a green pullover	on has	falls	over

'The girl who is wearing a green pullover falls over.'

c.

Dat	had	ze	niet	moeten	doen
that	had	she	not	must	do

'She shouldn't have done that.'

d.

Nadat	het meisje	gegeten	heeft	gaat	ze	spelen
after	the girl	eaten	has	goes	she	play

'After the girl has eaten, she goes to play.'

e.

De jongen	is	aan	het	fietsen
the boy	is	at	the	cycle

'The boy is cycling.'

f.

De jongen	zit	een boek	te	lezen
the boy	sits	a book	to	read

'The boy is (sat)/sits reading a book.'

Verbal density differs from the rate of subordination as calculated by Hunt (1970) and Verhoeven and Vermeer (1989) because it includes both finite and non-finite clauses (Loban 1976). It also differs from MLUw because it does not just measure length but arguably also depth, that is, the utterances illustrated in (4) demonstrate grammatical complexity at the clause level and this is different from simply stringing words together. This is particularly important for the L2ers who, as noted above, will as a result of their L1 be able to produce multi-word utterances from very early on. The second reason why verbal density is a suitable measure for present purposes is that it can be calculated irrespective of targetlikeness. If both finite and non-finite verbs are included in the count, it does not matter if they involve non-targetlike inflectional morphology. (The (non-)targetlike use of inflectional morphology is included in the proficiency score as part of the accuracy measure – see §4.6.)

Although it is almost inevitable that, like rate of subordination, verbal density will correlate with MLUw, for this measure to be considered a suitable indicator of morphosyntactic complexity, it should be the case that unlike MLUw, it does not develop significantly as a function of age in L1 children. As outlined above in §4.4.1.2, this is because this would result in a possible confound between the developing L1 and the developing L2 for the L2 children. Recall the logic behind this argument: if a measure of morphosyntactic complexity is observed to correlate with age for L1 children who are in the same age range (at time of testing) as L2 children, it is possible that L2 children will be limited by their developing L1 and this will consequently complicate the comparison both within the child L2 group, and between this group and the adult L2 group. Let us return to our example with the two L2 children, Dom and Helen. Both children have the same length of exposure but Helen is ten years old and Dom is seven. Let us assume that there is a significant age effect for verbal density in L1 children between the ages of five and ten. If Helen is observed to have a higher verbal density score than Dom, we cannot automatically conclude that she is more proficient in Dutch than Dom because her higher score may be attributable to her relative maturity rather than to her L2 development.

The verbal density measure was calculated for the three child L1 groups (7-, 9- and 11-year-olds) and for the native adults. All verbs were counted, including copula *zijn* ‘to be’ and the auxiliaries *hebben* ‘to have’ and *zijn* ‘to be’, *gaan* ‘to go’ and modals. It could be objected that including auxiliaries would unfairly favour subjects who chose to complete the task using periphrastic tenses, such as the present perfect (that is, the *voltooid tegenwoordige tijd*) and the future, over subjects who selected the (synthetic) present tense only. There are several reasons why this objection does not hold, however. Firstly, the use of a variety of tenses is an indication of grammatical sophistication, which is what verbal density is designed to measure. Secondly, there were no (L1) subjects who consistently used one tense only. Thirdly, excluding auxiliaries would mean that subjects who used the durative *is aan het* construction, as in (4)-e, would not be credited for this. This is particularly important for the L2ers as the use of this and the other durative construction illustrated in (4)-f is indicative of a certain level of grammatical sophistication in Dutch. Fourthly, excluding auxiliaries would mean that complex constructions which indicate grammatical complexity rather than simply choice of tense, such as the one given in (4)-d, would be excluded. Given the above, auxiliaries were included in the calculation.

The calculation of verbal density therefore involved all verbs. The results for L1 children and adults are presented in Table 12.

Table 12. L1 child and adult: Verbal density

Age group	N	Mean	Range	SD
7-year-olds	10	1.29	1.07-1.48	0.13
9-year-olds	10	1.41	1.24-1.65	0.12
11-year-olds	10	1.40	1.23-1.51	0.08
Adults	10	1.58	1.32-1.97	0.22

When all four groups were compared with each other, a statistically significant difference was observed (Kruskal-Wallis (data converted to ordinal scale): $\chi^2 = 12.273$, df = 3, $p < .01$).[23] The three child L1 groups do not significantly differ from each other, however (Kruskal-Wallis (data converted to ordinal scale): $\chi^2 = 4.846$, df = 2, $p > .05$; cf. the results for MLUw in §4.4.1.2.2). Consequently, although there is clearly some variation between the different child groups, they will, for present purposes, be considered to be one. As a group, the L1 children differ from the L1 adults. Before explaining how this difference is factored into the final proficiency score, one caveat concerning the native-speaker adult data is in order. The standard deviations given in Table 12 indicate that, as was observed for MLUw, the variation in the adult data is almost twice as much as in the child groups. This is less than ideal, but it is my opinion that when compared with the alternatives, the advantages of this measure over the alternatives outweigh this potential disadvantage.

The L1 children and adults who completed the picture description task differed in terms of verbal density. Assuming that this difference results from the differences in cognitive maturity between these two groups, this has serious implications for how this measure is implemented as part of the child L2 ~ adult L2 comparison carried out here. As noted above in the discussion of MLUw, it would be unreasonable to expect that an L2 child exceed the verbal density score achieved by an L1 child of the same age. Furthermore, if we assume that the findings for L1 Dutch children and adults can be extended to L1 English children and adults, it would also be unreasonable to expect an L2 child's interlanguage verbal density score to exceed that of her L1. Thus, it is highly likely that L2 children will intrinsically differ from L2 adults in terms of this measure. The verbal density scores for the L2 children and adults and for the L1 children and adults are presented in Table 13. Note that in this comparison, the status of an L2er as a child or adult must be defined in terms of age at time of testing. The most natural break in the data at which a division could be made is age 14, that is, when subjects are categorised according to age at time of testing, two clear groups are formed, with age 14 as the dividing line

[23] As with the MLUw data with these subjects, it was not possible to carry out an ANOVA with these data because they did not conform to the assumption that the variances within each group were equal, as shown by Levene's test of homogeneity of variances: Levene's statistic = 3.918, df1 = 3, df2 = 36, $p < .05$.

between the two.[24] Thus, for the purposes of the proficiency measure, all L2 subjects who are older than 14 years at time of testing are categorised as adults, and those younger than 14 years as children.[25]

Table 13. L1/L2 children and adults: Verbal density

Group	N	Mean	Range	SD
L1 children	30	1.37	1.07-1.65	0.12
L1 adults	10	1.58	1.32-1.97	0.22
L2 children	47	1.30	0.74-1.68	0.21
L2 adults	18	1.38	0.91-1.88	0.28

Both L2 groups pattern similarly with respect to their L1 age-equivalents. The mean for the L2 children is lower than the L1 children but the highest score is comparable, and a similar pattern obtains for the L2 adults when compared with the L1 adults. This suggests that, as expected, the adults (are able to) obtain higher scores. It might be objected that the adult L2ers are simply more proficient (better) in Dutch than the L2 children. This does not appear to be the case, however. Both the adult L2 and the child L2 data have a normal distribution (Shapiro-Wilk: $p > .05$ for both groups). Thus, it is not the case that the child L2 data are positively skewed and the adult L2 data negatively skewed such that this could explain the difference between the two groups in terms of highest scores. Note also that the maximum score achieved by the L2 children (1.68) is virtually identical to the L1 children (1.65). This further suggests that children as a group are limited by their cognitive immaturity relative to the adults.[26]

---

[24] This was most clearly the case for the picture description task data in the comprehension experiment (Chapter 6). In order to ensure comparability, the same age was used for the picture description task data in the production experiment, too (Chapter 5). Note that the data given in Table 13 are based on the comprehension picture description task only. This is because this is the task which the L1 children and L1 adults carried out.

[25] This means some subjects who are classed as L2 adults elsewhere (because they were exposed to Dutch at age eight or later) count as L2 children here (because they are younger than 14 years at time of testing). Whether these subjects should actually be classed as children rather than adults in all instances is discussed in Chapter 6 (§6.3.3.1.2 and §6.4.3.1.2).

[26] The difference between the L2 children and L2 adults is not significant (t-test (equal variances assumed): $F = 1.642$, $t = -1.374$, $df = 63$, $p > .05$), but this does not mean that the logic explicated here does not go through. The crucial point is that although the ranges of the L2 child and adult groups generally overlap, the scores at the top end of these ranges are higher for the L2 adults. The lack of a significant difference between the L2 children and the L2 adults may simply be because the adult group contains a range of proficiency levels and, consequently, as a group, the adults' average score will not necessarily be (close to) its maximum. This contrasts with the child L2 group: although this group also contains a range of proficiency levels, the maximum which these subjects can reach is argued to be constrained as a result of their younger age. The observation that both sets of data have a normal distribution supports this claim.

What the intrinsic differences between children and adults in terms of potential verbal density score mean is that L2 adults who score higher on this measure should not automatically be regarded as more proficient than L2 children whose scores are slightly lower. Such a disparity might stem from the inherent difference in cognitive maturity between the child and adult subjects rather than from a contrast in their interlanguage grammars. In order to ensure that this age factor does not adversely influence the L2 subjects' proficiency scores, the scores for this variable (and the lexical diversity measure – see §4.5) are converted into standardised scores for the child and adult groups separately, before they are combined into the final proficiency score (cf. §4.7). A brief summary of the rationale behind standardised scores should serve to illustrate why this is necessary.

Standardised scores, or *z-scores*, allow two or more scores on the same variable to be compared with each other by taking into account the position of the scores within the distribution of the samples from which they are taken. For example, imagine two language learners take a Dutch language test. Person A takes Test I and scores 48%, whereas person B takes Test II and scores 60%. Assuming that the tests are similar enough for us to be able to consider that they test the same thing, can we conclude that person B is more proficient in Dutch than person A? In order to be able to answer this question, we need to know how the other people who took these two tests performed and compare our two learners' scores with their scores. If, for example, it turns out that the average score on Test I was 42% and on Test II, it was much higher at 65%, person A should actually be regarded as more proficient than person B, because she scores higher than the average score for Test I, whereas person B's score is below average on Test II. This situation is comparable to our child ~ adult comparison on the verbal density measure: the adults generally score higher than the children, and hence, in order to make a fair comparison between these two groups, it is first necessary to evaluate each subject's score relative to the rest of the group. Z-scores do just that. They indicate how many standard deviations a given point differs from the sample mean. Negative and positive z-scores are below and above the average, respectively. The verbal density score for each L2 subject was thus converted to a z-score for the L2 children and L2 adults (as determined by age at time of testing) separately and it is these z-scores which are used in the overall proficiency score. Both the raw verbal density scores and the z-scores for each subject are provided in Appendix B.

This section outlined why and how verbal density is used as an indicator of grammatical/morphosyntactic complexity in the present study. An analysis of comparable child L1 and adult L1 data demonstrated that verbal density did not differentiate between the child L1 groups but it did differentiate the L1 children from the L1 adults. In order to factor the potential effect this difference might have into the proficiency score, the verbal density scores for the two L2 groups were converted into z-scores within the child L2 and adult L2 groups separately, before being combined with the results of the other two measures to derive a single proficiency score.

4.4.5 Summary

This section presented and evaluated four measures of morphosyntactic complexity: MLU, rate of subordination, rate of verbal utterances, and verbal density. Data from L1 children and adults collected using exactly the same picture description task as the L2 subjects were used to determine whether MLUw and verbal density were suitable for use in a child L2 ~ adult L2 comparative study. Provided certain caveats are borne in mind, this latter measure was considered the most suitable. It was found that the L1 children and L1 adults differed from each other on both of these measures and conceptual and empirical arguments were put forward to show that this age-related effect is also present in the L2 data. Consequently, in order to factor out any influence this age effect might have, the L2 subjects' scores were standardised within each age group (child vs. adult at time of testing) as a preliminary step prior to the final proficiency score being calculated.

Given our definition of proficiency as 'the ability to produce lexically, morphologically and syntactically complex and accurate utterances in the TL', it is also necessary to incorporate a measure of lexical complexity. This is the focus of the following section.

4.5 Measuring lexical complexity

In the context of the present study, lexical complexity is understood as synonymous with lexical diversity. The range of lexical items which a subject uses in completing the picture description task will be taken as an approximate indication of the diversity present in the learner's developing lexicon. Lexical diversity (or richness) is traditionally measured using the 'Type/Token Ratio' (TTR). There is, however, evidence to suggest that this measure is inadequate. This section briefly reviews some of this evidence before outlining the measure which will be adopted here, namely 'Guiraud's index' ($V/\sqrt{N}$). As for verbal density (and MLUw), the child L1 and adult L1 data will be used to determine the extent to which there is an age effect in this domain. It will be observed that in the context of the picture description task employed here, there is no difference between the three child L1 age groups but the L1 children (as one group) do (unsurprisingly) differ from the L1 adults. As outlined in the previous section, the implications of this observation for the child L2 ~ adult L2 comparison will be factored into the proficiency score by standardising the child L2 and adult L2 scores within each group separately before combining them with the other two measures to derive a single proficiency score.

The traditional Type/Token Ratio is calculated by dividing the number of types (V) in a sample by the number of tokens (N). For example, five different types in a sample of 20 tokens would result in a TTR of .25 (5/20), whereas if ten different types were produced the resulting TTR would be .5 (10/20). A TTR of .5 is assumed to reflect a more diverse lexicon (or larger vocabulary) than a TTR of .25. The assumption is that this score provides an index of lexical development independent of sample size and consequently it is widely used in both L1 and L2 acquisition studies.

The TTR has, however, been shown to be inadequate in several ways. Richards (1987) shows that TTR *is* affected by sample size. TTR is artificially deflated as a result of an increase in tokens in a sample. Imagine that at time I, a child produces five different types in a sample of 20 tokens. At time II, the child is slightly older and is more talkative. She produces the same five types as at time I, but her sample consists of 30 tokens. Her TTR would actually suggest that her lexicon was less diverse at time II (.17) than at time I (.25), but this is not the case. One possible solution to this problem, Richards suggests, is to standardise the number of tokens over which the TTR is to be calculated, although just how many such a standard number should be is left for future research.

Vermeer (1992), reporting on work by Van Hout and Vermeer (1988), shows that in addition to sample size, TTR is also sensitive to the stage of acquisition being measured and the way in which the relationship between types and tokens develops with increasing proficiency. For example, if there is a linear relationship between types and tokens as proficiency increases, that is, if the number of types increases at the same rate as the number of tokens, then the TTR will remain constant and no development will be measured. If, on the other hand, the number of tokens increases more rapidly than the number of types, for example at the point at which determiners are acquired, there will be a slight dip in the TTR, suggesting lack of development (or even backsliding) whereas in reality, language learners at this stage will have made considerable steps in their linguistic abilities. This is akin to the dip found above when more tokens are produced, which is essentially what, for example, the acquisition of determiners entails.

In their comparison of spontaneous data measures and the standardised *TAK* test, Verhoeven and Vermeer (1989) use the (absolute) number of different words as a measure of lexical diversity. In the *TAK*, lexical proficiency is assessed using passive and active vocabulary tasks. In the passive vocabulary task, subjects are presented with four pictures and asked to point to the one the experimenter names. In the active vocabulary task, subjects are asked questions about pictures or they are asked to complete sentences using information given to them on a picture. In the spontaneous data, which, recall, included a fixed number of utterances (200 per child), lexical proficiency was measured by simply counting the total number of different words. This was found to correlate moderately with the two vocabulary measures from the *TAK* ($r = .55$, $p < .01$ for the passive vocabulary task and $r = .66$, $p < .01$ for the active vocabulary task). Although this finding might suggest that number of different words could be used here, the fixed number of utterances of 200 is unlikely to be achieved for most of my (L1 or L2) subjects. Furthermore, if the number of utterances needs to be standardised, it is highly likely that this number will have to be unreasonably low if it is to be the same for all subjects (given that some subjects produce much less than others) and this would render any calculations virtually meaningless.

Broeder, Extra and van Hout (1993) evaluate several different measures of lexical richness, including TTR, using data from 20 adult L2ers with various L1/TL combinations. They consider a measure to be successful if it demonstrates progress in vocabulary for at least some of their subjects over the 30-month observation period. They reject the TTR because it shows a decrease over time. The measure known as Guiraud's index or the 'Indice de Richesse' ($V/\sqrt{N}$), on the other hand, shows clear development over time. As the authors note, this finding is supported by the results from Vermeer (1986), who found a strong correlation between Guiraud's index and a standardised vocabulary test, as well as other proficiency tests. Guiraud's index is calculated by dividing the number of types (V) by the square root of the number of tokens. By taking the square root of the number of tokens, the problem of a negative correlation with increasing sample size, as described above for TTR, is obviated.

Richards (1987) notes, however, that Guiraud's index is *positively* correlated with sample size. He calculates Guiraud's index for cumulatively larger samples taken from the transcript of a 2-year-old child and finds that when the sample contains fewer than 250 tokens, any sharp increase in the next sample in the number of tokens results in a concomitant increase in Guiraud's index. Richards (1987:209) suggests, therefore, that if both TTR and Guiraud's index are susceptible to fluctuations in sample size, once sample size is standardised, there should be relatively little difference between the two measures. Broeder *et al.* (1993:154) acknowledge this problem but argue that given that increased output generally accompanies increasing proficiency, a correlation between sample size and Guiraud's index 'is not by definition objectionable'.

In order to assess whether Guiraud's index would be appropriate for present purposes, it should determined whether the subjects' samples contain enough tokens, i.e. 250 or more. To this end, the number of tokens in each of the 107 L2 samples (for both the production and comprehension experiments) and the 40 L1 samples was counted (using the FREQ command in the CLAN program). Two of the L2 samples were excluded because they contained very few tokens (just two and four, respectively). All of the L1 adult samples contained more than 250 tokens, as did 21 of the 30 child L1 samples. Of the remaining nine child L1 samples, eight had more than 200 tokens and the remaining one had more than 150. Around three quarters (78/105) of the L2 samples contained more than 250 tokens and 91 of all 105 L2 samples contained more than 200 tokens. Only seven L2 samples contained less than 150 tokens. It would seem reasonable to conclude therefore that the 'sample size effect' which Richards (1987) mentions is not relevant for most of the samples in my data set. Consequently, the lexical component of the proficiency score will be based on Guiraud's index, i.e. $V/\sqrt{N}$.[27]

Similar to our discussion of MLUw and verbal density, in order to assess the suitability of Guiraud's index as an indicator of lexical complexity, it is necessary to determine the extent to which native-speaker children and adults vary on this variable (especially given that vocabulary size is known to vary across speakers). To this end, Guiraud's index was calculated for the L1 children and adults. The results are presented in Table 14.

Table 14. L1 children and adults: Lexical diversity

Age group	N	Mean	Range	SD
7-year-olds	10	6.04	5.01-7.44	0.72
9-year-olds	10	6.27	4.97-7.30	0.59
11-year-olds	10	5.80	5.25-6.14	0.74
Adults	10	7.07	6.07-8.22	0.74

A one-way ANOVA indicates that there is a significant difference between groups ($f = 8.161$, $p < .001$). Post-hoc Bonferroni tests show that the child groups do not differ from each other but that they each differ from the adults (7-year-olds vs. adults, $p < .01$; 9-year-olds vs. adults, $p < .05$; 11-year-olds vs. adults, $p < .001$). Note that the claim regarding the lack of difference between these three child L1 groups is only intended to hold for this picture description task: in this context, these three groups produce a similar range of lexical items, one which is significantly more restricted than the range of items produced by adults on the same task. I wish to make no claims regarding L1 vocabulary development in general.

In the previous section on verbal density, it was argued that given the difference between L1 children and L1 adults, on the one hand, and the lack of difference between the child L1 groups, on the other, the L2 subjects' scores should be evaluated separately for children and adults, prior to being combined with the other components of the proficiency score. I propose to use the same logic here. On verbal density, differences were observed between the L2 children and adults with respect to the maximum scores these subjects obtained. These data paralleled the L1 data. Such a pattern is also found for the

[27] Vermeer (2000) proposes a different measure which incorporates the degree of difficulty of the words used, arguing that this is more effective than, for example, Guiraud's measure, when vocabulary reaches 3000 words or more. The use of such a measure is beyond the scope of this thesis.

lexical diversity measure. Table 15 presents the data for both L1 and L2 children and adults.

Table 15. L1/L2 children and adults: Lexical diversity

Group	N	Mean	Range	SD
L1 children	30	6.04	4.97-7.44	0.58
L1 adults	10	7.07	6.07-8.22	0.74
L2 children	47	5.34	3.58-7.54	1.08
L2 adults	18	6.89	3.96-8.94	1.55

Here, we find the same pattern as in Table 13 above on verbal density. Both L2 groups pattern similarly to their L1 age-equivalents. The mean for the L2 children is lower than for the L1 children but the highest score is comparable, and a similar pattern obtains for the L2 adults when compared with the L1 adults. (Some of the L2 adults actually score higher than the highest L1 adults on this measure.) Hence, similar to verbal density, the lexical diversity scores will also first be converted to z-scores for each group prior to use in the general proficiency score. The individual lexical diversity scores and z-scores are given in Appendix B.

To summarise: For lexical complexity, three different measures were evaluated: type/token ratio, total number of different words, and Guiraud's index. It was the latter measure which was selected for use here. Like verbal density, scores on Guiraud's index were not found to significantly differ amongst L1 children but they did differ between the L1 children and L1 adults. The same conceptual and empirical arguments as in the previous section were put forward to show that this age-related effect was also present in the L2 data. For this reason, the L2ers' lexical diversity scores are converted to z-scores before being combined into the final proficiency score. Before this is possible, however, the third component of the proficiency score, that is, the accuracy measure, is presented in more detail.

4.6 Measuring accuracy

An assessment of L2 proficiency made on the basis of complexity alone would be insufficient because complexity closely interacts with accuracy (Lalleman 1986). As noted above (§4.3), L2ers who produce complex yet inaccurate utterances should not be considered more proficient than L2ers who produce less complex but more accurate utterances. In order to ensure that the interaction between these two factors is taken into account, the proficiency score incorporates a measure of accuracy as well as a measure of complexity. As discussed in §4.3, Whong-Barr and Schwartz (2002), following Larsen-Freeman (1983), used the percentage of errorfree utterances for this purpose. In this section, we discuss the evidence for the usefulness of this measure

(§4.6.1), before outlining how exactly it is implemented in the present study in §4.6.2 through §4.6.4.[28]

4.6.1 Rate of errorfree utterances

The rationale behind this calculation is simple: the more developed/advanced L2ers are, the less likely they are to commit errors.

As a measure of accuracy, Larsen-Freeman and Strom (1971) choose the rate of errorfree utterances instead of the proportion or number of errors for several reasons. Firstly, they argue that in calculating the number of errors, it is not clear how the different errors should be weighted, that is, whether certain errors should be considered more serious than others. Also, given that some subjects produce more utterances than others, simply counting the absolute number of errors would be inappropriate. Second, assessing proficiency using number of errors implicitly assumes linear development in language acquisition and this has been found to not always be the case. On their way to acquiring a given phenomena, L2ers often progress through a series of peaks and troughs, and consequently any assessment of L2 proficiency based on the commission of errors (alone) would skirt over such changes.

Note, however, that these criticisms made of an error analysis could also be levelled at an analysis based on rate of errorfree utterances alone. The relative weighting of different errors is also a consideration in deciding whether a given utterance is errorfree or not. Certain errors may be considered too minor to render a whole utterance non-targetlike, for example, and others may be ignored. In addition, assessing L2 proficiency in terms of errorfree

[28] A preliminary version of the proficiency score (reported in Unsworth 2004a; 2004b) included a secondary accuracy variable, namely the proportion of utterances where the subject did not switch to English. This secondary (somewhat *ad hoc*) measure was motivated by the observation that the less advanced subjects seemed to resort to using more English lexical items, whereas the more advanced subjects did not, and so rather than exclude all utterances containing L1 items (which in some cases would have resulted in very few useable utterances), I wanted to incorporate these non-targetlike productions into the final score in some other way. It has since come to my attention, however, that this measure is inappropriate for a number of reasons. Firstly, code-switching was not entirely inappropriate in the context in which my data are collected because both a native speaker of English (me) and a native speaker of Dutch (student assistant) were present. Although the whole session was conducted in Dutch (as far as was possible), many of the participants knew that I am a native speaker of English (or at least that I speak English) and hence they were potentially in a (semi) bilingual mode (Grosjean 1998), that is, both of their languages were active. Consequently, penalising them for using English words would be rather unfair. Secondly, it is difficult to distinguish genuine code-switching (legitimised by my presence) and the use of English words because of a lack of knowledge of the equivalent Dutch words. The situation is further complicated by the fact that native speakers of Dutch often insert English words into their utterances, too. Finally, the extent to which a bilingual speaker code-switches has been shown not to be related to proficiency (see Cantone and Müller 2004 on early 2L1A development).

utterances also implicitly assumes linear development in language acquisition, namely that as proficiency increases, so does the rate of errorfree utterances. In an attempt to address some of these criticisms, the present study is very explicit about what should count as an error. This contrasts with Larsen-Freeman (1983:290), who fails to give any specific criteria (or examples), simply stating that 'errors involving prepositions, articles and verb tense were most prevalent'. The remainder of this section presents the specific criteria for determining errors which are used here.

4.6.2 Deciding what counts as an error and which errors should be ignored

As suggested above, all errors are not equal. In this section, I first discuss some of the problems involved in determining what should or should not count as an error before outlining the criteria which I have used to code my data.

When utterances include partial repetitions and/or self-corrections, they are evaluated on the basis of the repeated/corrected form (the part which is not between angle brackets in the transcription) because I assume that this is the form which the subject intended to count as the final form in much the same way as native speakers correct themselves.

When subjects make repeated errors, for example, consistently using a particular word incorrectly or systematically using a non-targetlike verb form, we might consider only counting the first instance of such an error as an error so as not to overpunish them (which would potentially artificially deflate their score as a result of one persistent error). However, the variability which is often pervasive in L2 grammars, especially with regard to word order and inflectional morphology, means that subjects will sometimes produce a certain error and at other times they will be targetlike. Consequently, if an error were initially counted as such, the subject in question would not be rewarded when s/he does produce a targetlike form, but if the error were ignored, the L2er would be attributed with more knowledge than s/he has. Hence, (morphosyntactic) errors which are repeated are counted separately.

Nevertheless, there are certain errors which are so pervasive, namely errors of grammatical gender, that there is definitely a case for ignoring them completely. Given that even the most highly proficient subjects commit such errors, incorporating them into the proficiency score would be rather pointless as they fail to differentiate between subjects.[29] Some gender-related examples, all of which are ignored in the accuracy measure, are given in (5). Dutch has two genders, non-neuter and neuter. The definite determiner for non-neuter nouns is *de* and for neuter nouns it is *het*. (For more details, see the discussion of Weerman *et al.* 2003 in §1.4.3.3.)

[29] This approach was also taken by Tran (2005a; 2005b), who discounted case errors in her L2 German subjects for similar reasons.

(5) a. non-targetlike determiners (*de* used instead of *het* and vice versa)

*SUB:	de poes	kijk	naar	**de** eendje.
	the cat	look	to	the duckling

'The cat is looking at the duckling.'

b. non-targetlike adjectival inflection (ignored because it is often not clear what gender the subject attributes to a given noun)

*SUB:	en die	heeft	een zwembad,	een **kleine** zwembad.
	and that	has	a swimmingpool	a little swimmingpool

'And she's got a swimming pool, a little swimming pool.'

c. non-targetlike pronouns (except when the pronoun clearly refers to an animate object for which the grammatical gender is clear)

*SUB:	en hij	wilt	**hem**	pakken.
	and he	wants	him	take

'And he wants to take it.' (where *hem* 'him' refers to *het koekje* 'the cookie')

Other 'errors' which are ignored involve the use of accepted dialectal variants.[30] Several (child L2) subjects use the 3PPL possessive pronoun, *hun*, as a subject pronoun. Although considered non-standard, this is a relatively common feature of spoken Dutch (which if anything, shows the subjects' proficiency rather than lack of it) and as such, this is not counted as an error. An example is given in (6).

(6) *SUB:	en	hij [//]	hun	allebei	vallen.
	and	he	them	both	fall

'And them [=they] both fall.'

Several subjects also use masculine subject and possessive pronouns to refer to females, as illustrated in (6).

(7) *SUB:	dan	moet	ze [= het meisje]	met	z(ij)n fiets	lopen.
	then	must	she [= the girl]	with	his bike	walk

'And then she has to walk with his [= her] bike.'

This is a dialectal feature common to the southern province of Brabant (J. Swanenberg, p.c. 14 September 2004, Weijnen 1937), where several of the subjects live. For these subjects, this non-standard feature was not classed as an error.

There were also several other types of non-targetlike lexical items which were also not counted as errors. These include L1 or self-created target-language

[30] Native-speaker adults were used as an approximate guideline here.

nouns, verbs, adverbs and prepositions. Using English nouns, verbs, adverbs and prepositions instead of Dutch ones or creating target language forms which do not exist is assumed to reflect a lack of target language vocabulary rather than a lack of grammatical knowledge, whereas using English function words instead of their Dutch equivalents or using Dutch function words in a non-targetlike fashion can be taken as an indication of non-targetlike grammar.[31, 32] The consequences of not counting these non-targetlike productions as errors are as follows. Utterances which are syntactically and

---

[31] The only instance where this might not be the case is with prepositions which mark grammatical relations, such as *van* 'from/of'. In all 105 files, there were, however, only four utterances where English 'of' was used in this way. Two were produced in utterances which were excluded because more than half of the words were in English (see below) and the other two, produced by two L2 children, are given in (i). There were no such uses of 'from'.

(i)	a.	en	de meisje	verft	een picture	of een paard	en	een man
		and	a girl	paints	a picture	of a horse	and	a man
		'And the girl paints a picture of the horse and a man.'						
	b.	ze	komt	uit	of de huis	en naar	de dak	
		she	comes	out	of the house	and to	the roof	
		'She comes out of the house and (goes) to the roof.'						

[32] When subjects used an L1 verb, they varied as to whether they inflected this with Dutch morphology:

(i)	*SUB:	en	zij	take	een appel.
		and	she	take	an apple
		'And she takes an apple.'			
(ii)	*SUB:	en	then	zij	swIm6@u [= zwemmen].
		and	then	she	swim$_{PL}$
		'And then they are swimming.'			

Using an L1 verb with TL morphology clearly demonstrates missing lexical knowledge. It is not clear, however, whether using an L1 verb without TL morphology demonstrates missing morpho-syntactic knowledge as well as missing lexical knowledge. Rather, this might reflect the extent to which a given subject code-switches (or is able to code-switch). Given that the proficiency measure is not intended to measure this aspect of multilingual language knowledge/use, the latter type of utterance are classified as errorfree if all other relevant aspects are errorfree (according to the criteria set out here). Further motivation for coding these utterances in this way is the fact that in many cases, where the L1 verb ends in a phoneme which is homophonous to the appropriate TL morphology, it would be impossible to determine whether the inflectional morphology was present. For example, in (iii), the L1 verb ends with /t/ and the targetlike morphology is /t/.

(iii)	*SUB:	en	ze	put	lucht	in het.
		and	she	put	air	in it
		'And she puts air in it.'				

morphologically errorfree, but which contain a non-TL verb, adverb, noun or preposition are counted as errorfree. This means that subjects are not 'punished' because they do not know a word when that word is in an utterance which is otherwise targetlike. The non-targetlike aspect of such utterances is factored into the proficiency score via the lexical richness measure (where these items are counted as tokens but not as types, because they belong to the L1).[33] If these non-TL words were classified as errors and the utterances in which they occur were counted as non-targetlike, it would mean that this 'error' would be counted twice (once in the accuracy measure and once in the complexity measure) and this would unfairly deflate a subject's proficiency score. Note that the same reasoning cannot be used for TL items used in a non-targetlike way, which, as noted below, are classed as errors. If these items are included in the lexical complexity measure and *not* counted as errors, their non-targetlike use would not be included in the proficiency score. Rather, subjects who produced such items would be positively rewarded for having more TL lexical items in their vocabulary even though these items may be used in a non-targetlike fashion. By counting them as errors, we are able to capture this non-targetlike aspect of their use.

As stated in the introduction, all phonological/pronunciation errors are ignored. When a given form is ambiguous between an L1 and an TL item, the subject is always given the benefit of the doubt and the form in question is thus classified as belonging to the TL. An example of such an item is *in*. Also, subjects sometimes use the form *één* instead of *een*. This is not counted as an error because it is impossible to tell whether this is a lexical or phonological error.

As noted in §4.3.1, elliptical answers in response to questions are excluded because they may add unwanted variability to the proficiency score. Targetlike utterances involving elided or missing constituents, as in topic-drop or subject-drop in main clauses, are however included in the analysis. An example of topic-drop is given in (8). As long as the remainder of the utterance does not contain any of the errors outlined in §4.6.4, the use of topic-drop and subject-drop is counted as targetlike because native speakers produced similar utterances.

(8) *SUB: gaat ze stiekem lezen in bed.
goes she secretly read in bed
'She goes and reads in bed secretly.'

One final 'error' which was excluded is the use of the past or present perfect (*voltooid tegenwoordige tijd*) with the connective *dan* 'then', as in (9). In Dutch,

---

[33] Likewise, when TL words were provided by the subject's interlocutor during the task (usually at the request of the subject), these words are included in the assessment of the accuracy of the utterance in which they are contained, but for the lexical diversity measure, they are counted as tokens but not as types.

there are two connectives used for English 'then'. Generally, *dan* is used with present tense and *toen* with past or present perfect tense. Many of the L2 subjects used *dan* with the past tense, however. Although this might be the result of L1 transfer, such utterances were not counted as errors because the native speakers also produced such utterances.

(9) *SUB: dan ging hij weg
then went he away
'Then he went away.'

This contrasts with the use of the present tense with *toen*, however. Native speakers rarely produce such utterances and hence, this was counted as an error.[34]

4.6.3 Utterances which are excluded

The following types of utterance are excluded for the verbal density and rate of errorfree calculations. Examples are given in CHAT format, including the relevant 'postcodes' (in square brackets), which mark the utterances as belonging to a given category.[35]

(10) One-word utterances consisting of a proper name

*SUB: Paddington+Beer. [+excl]
Paddington Bear
'Paddington Bear.'

(11) One-word utterances consisting of *ja* 'yes' or *nee* 'no', including in response to a yes-no question

*EXP: ze bakken.
they bake
'They're baking'

[34] Further evidence for this difference between *toen* and *dan* comes from an internet search using GOOGLE. A search for *toen* and *dan* in combination with the present and past tense forms of the highly frequent verbs *gaan* 'to go' and *doen* 'to do' delivered the following results. *Toen* occurred with the past tense form of *gaan* 222 times more frequently than with the present tense form, and with the past tense of *doen* 101 times more frequently than the present tense form. This contrasts with *dan*, which occurred with the present tense form just seven times more frequently than the past tense forms of both *gaan* and *doen*. In other words, *dan* is used with the past tense forms of *gaan* and *toen* significantly more often than *toen* is used with the present tense forms.

[35] Postcodes, which always appear after the utterance delimiter, are used in CLAN/CHAT to mark an utterance as belonging to a given category. Utterances coded in this way can subsequently be selected using the KWAL function or they can be excluded from a particular operation by placing the relevant postcode(s) in a file and excluding this file by using the code –s@*filename*.

*SUB: ja. [+ excl]
yes
'Yes.'

(12) Utterances which are incomplete (marked as such with postcode [+ inc]), due to:

a. self-interruption

*SUB: en hij heeft +//. [+ inc]
and he has
'And he has …'

b. interruption by interlocutor[36]

*SUB: de volgende is een +/. [+ inc]
the next one is a
'The next one is a …'

*EXP: Kijk je nooit naar sesamstraat?
look you never to sesame street
'Do you never watch Sesame Street?'

*SUB: ja. [+ excl]
yes
'Yes.'

c. trailing off

*SUB: Grover met <his> [//] zijn vriend +… [+ inc]
Grover with his friend
'Grover with his friend…'

(13) Utterances where the subject completes the interlocutor's utterance (marked with [+ excl])

*EXP: en dan ze is +…
and then she is
'And then she is …'

*SUB: ++ boos. [+ excl]
angry
'… angry.'

[36] This was very infrequent.

(14) Utterances containing unidentifiable material, as indicated by either xx or xxx

*SUB: en de dame bij wie xxx is een taart of zo.
and the lady at who is a cake or so

aan het bakken.
at the bake
'And the lady at whose house xxx is baking a cake or something at.'

(15) Formulaic utterances, such as *weet ik niet* 'don't know', *sorry* 'sorry', *even kijken* 'let's take a look', *doe maar* 'go on then', *wacht even* 'wait a minute', *oh*, *oke* 'okay', *wat is dat?* 'what's that?' (marked as such with postcode [+ form])

(16) Exact repetition of interlocutor's utterance or subject's own utterance (marked as such with postcode [+ R])

*EXP: niet blij # nee.
not happy no
'(He's) not happy, no.'

*SUB: niet blij. [+ R]
not happy

(17) Utterances where more than half of the words (counted in words) are in English (marked as such with postcode [+ eng])

a. *SUB: I don't know how to describe it. [+ eng]

b. *SUB: <dan zij (is)> [//] zij [//] but she pretends
then she is she

she is <sleep> [//] slapen [+ eng]
she is sleep
'But she pretends she is sleeping.'

(18) Utterances which contain scrambled/non-scrambled objects (marked as [+ scr])

*SUB: ik niet lezen de boeken yet. [+ scr]
I not read the books
'I haven't read the books yet.'

Utterances containing (non-)scrambled objects were excluded to ensure that the proficiency measure did not measure exactly the same as the experimental tasks.[37]

[37] In her study of Root Infinitives in child L2 acquisition (see Chapter 1), Tran (2005a; 2005b) also used a picture description task based on Whong-Barr and Schwartz (2002) to measure L2 proficiency. Errors in verbal inflection were included in her measure of accuracy leading to a potential confound between the topic of investigation in the experimental tasks in her study,

4.6.4 Criteria for an errorfree utterance

In order to be categorised as errorfree, an utterance must not contain the morphological, syntactic and lexical errors listed below in §4.6.4.1 through §4.6.4.3. Such utterances were marked as errorfree with the postcode [+ EF]. Note that when an utterance contains a word marked with [?] (indicating best guess), it is assumed that this best guess is indeed correct, that is, the utterance is judged with this word, unless the word in question is crucial to determining whether the utterance is errorfree, for example, if it is not clear if the verb is inflected. In this latter case, the utterance is excluded.

4.6.4.1 Morphological errors

Morphological errors include:

(19) Subject-verb agreement

*SUB: de meisje **slapen**.

Target: het meisje **slaapt**.
the girl sleeps
'The girl is sleeping.'

(20) Non-targetlike form of verb stem (present, preterite or past participle)

a. *SUB: de jongen **hulpt** de eendje.

Target: de jongen **helpt** het eendje.
the boy helps the duckling
'The boy is helping the duckling.'

b. *SUB: een moeder heeft een nieuwe zwembad **gekopen**.

Target: een moeder heeft een nieuwe zwembad **gekocht**.
a mother has a new swimming pool bought
'The mother has bought a new swimming pool .'

(21) Non-targetlike form of noun[38]

*SUB: de meisje rijden het **fietsen**.

---

namely verbal inflection, and the proficiency score. Tran has however since recalculated the rate of errorfree utterances for her subjects excluding verbal inflection errors, and once again, she found a correlation between proficiency and lack of RIs (B.D. Schwartz, p.c., 5 August 2005).

[38] The non-targetlike form of the noun is clearly not the only error in this utterance. Given that the focus here is on errorfree utterances, separate errors were not coded individually.

Target:	het meisje	rijdt	op de **fiets**.
	the girl	rides	on the bike

'The girl is riding the bike.'

4.6.4.2 Syntactic errors

Syntactic errors include:

(22) Non-targetlike word order, including failure to use V2 in matrix clauses and non-final verb in subordinate clauses

a. *SUB: hier de jongen **wil** de vliegel in z(ij)n hand doen

Target:	hier	**wil**	de jongen	de vogel
	here	wants	the boy	the bird
	in	z(ij)n	hand	doen
	in	his	hand	do

'Here the boy wants to put the bird in his hand.'

b. *SUB: omdat ze **wil** de boek lezen.

Target:	omdat	ze	het boek	**wil**	lezen
	because	she	the book	wants	read

'Because she wants to read the book.'

(23) Failure to split up particle and verb in particle verbs

*SUB: en dan die andere jongen **aanraakt** de andere jongen

Target:	en	dan	die	andere	jongen
	and	then	the	other	boy
	raakt	die	andere	jongen	**aan**
	touches	the	other	boy	on

'And the other boy touches the other boy.'

(24) Missing functional elements, including determiners, complementisers and copula/auxiliary verbs

a. *SUB: en zij weet zij heb een appel G6Ge:t@u [=gegeten].

Target:	en zij	weet	**dat**	zij	een appel
	and she	knows	that	she	an apple
	gegeten	heeft			
	eaten	has			

'And she knows that she has eaten an apple.'

b.	*SUB:	en	die meisje	boos
	Target:	en	het meisje	**is**	boos
		and	the girl	is	angry

'And the girl is angry.'

(25) Missing *er* (e.g. referring to previously mentioned object)

*SUB: en dan de meisje doet water in.

Target:	en dan	doet	**er**	het meisje	water	in.
	and then	does	it	the girl	water	in

'And then the girl puts water in it.'

4.6.4.3 Lexical errors

Lexical errors include the non-targetlike use of targetlike forms, that is, words which exist in Dutch but which are used with a non-targetlike meaning or function, such as:

(26) Non-targetlike preposition

*SUB: en dan fietst de oma **in** de steen

Target:	en dan	fietst	de oma	**tegen**	de steen
	and then	cycles	the grandma	against	the stone

'And then the grandma cycles into the stone.'

(27) Non-targetlike subordinating coordinator

*SUB: en dan doet ze net **als** ze in slaap is gevallen

Target:	en dan	doet	ze	net
	and then	does	she	just

alsof	ze	in slaap	is	gevallen
as if	she	in sleep	is	fallen

'And then she pretends that she's fallen asleep.'

(28) Non-targetlike collocation

*SUB: haar tanden **wassen**.

Target:	haar tanden	**poetsen**.
	her teeth	clean

'Cleaning her teeth.'

(29) Target language word used with non-targetlike meaning (often as a result of L1 influence)

*SUB:	dan	de jongen	**trapt** [= trips]	
Target:	dan	**valt**	de jongen	**om**
	then	falls	the boy	over
	'And then the boy falls over.'			

Lexical errors were only counted as an error on first occurrence. When the same error was made a second time and the remainder of the utterance in which it occurred was targetlike, the utterance was classed as errorfree. This was because it became clear during coding that the lexical errors were generally very consistent and hence the argument put forward above regarding repeated morphosyntactic errors being variable does not hold here.

Data were coded and checked by a native speaker (student assistant) and near-native speaker (the present author). An example transcript is provided in Appendix B. To ensure that the error-coding procedure was employed consistently, inter-coder reliability was calculated using the agreement index, Cohen's Kappa. This is used to measure the extent to which two (or more) coders' judgements agree with each other, while taking into account the likelihood that any agreement would be due to chance. A sample of 10% of the 107 transcripts were randomly selected and Cohen's Kappa was found to have a value of 0.829 ($p < .001$ (approx.)). Any value between 0.81 and 1.00 is taken to indicate almost perfect agreement (Landis and Koch 1977). We can therefore assume that the error-coding procedure was consistently employed by the two judges.

4.6.5 Calculating the rate of errorfree utterances

The rate of errorfree utterances was calculated by dividing the number of errorfree utterances by the total number of utterances (excluding those listed in §4.6.3). Comparing this set of L2ers' scores with those of native-speaker controls is different from the situations discussed previously in regard to verbal density and lexical diversity because here the adult native speakers perform at ceiling: the average score for the ten adults, using the criteria outlined above, is 99.7%. This means that the L2ers' proficiency cannot be evaluated by comparing their errorfree scores with the (range and variability) of the natives'. The L2ers' errorfree scores must be compared with each other. This will be possible once the errorfree scores are combined with the verbal density and lexical diversity scores. This is achieved using a 'principal components analysis', the details of which are presented in the following section.

4.7 Computing a single score

The proficiency measure consists of three sub-scores: a morphosyntactic complexity score (verbal density), a lexical complexity score (Guiraud's index) and an accuracy score (the percentage of errorfree utterances). Converting these three sub-scores into a proficiency score requires reducing them to a single factor. This is achieved using a principal components (or factor) analysis. In the remainder of this section, I briefly outline how principal components analysis works and how it is employed in the present study. Only the most relevant parts of this statistical method are discussed. For an overview of how principal components analysis works, see for example StatSoft (2004).

Principal components analysis is a means of reducing the number of variables in a data set, and of detecting structure between these variables. This is achieved by modelling the data on a three-dimensional scatterplot to obtain one or more new variables (or components) which account for as much variance amongst the original variables as possible. It is assumed that the original variables correlate, that is, that they measure the same construct. In our case, it is assumed that they all measure some aspect of L2 proficiency. These variables are not identical, however, because if they were, two would be redundant. Principal components analysis is thus a means of extracting the commonalities between several variables in such a way that as much variance as possible amongst these variables is accounted for by the resulting components or factors.

Two principal components analyses were carried out using the statistics package SPSS: one for the picture description task which was carried out as part of the production experiment in Chapter 5 and one for the picture description task which was carried out as part of the comprehension experiment in Chapter 6. Given that it is the L2 subjects only who will be assigned proficiency scores, these analyses excluded the L1 data used earlier in the chapter. For both tasks, it was first established whether the assumption holds that the three variables correlate with each other. As the correlation matrices given below indicate, this was found to be the case. Note that as discussed in §4.4.4 and §4.5, the scores for verbal density and lexical diversity are z-scores calculated for the adult and child (at time of testing) groups separately.

Table 16. Proficiency measure (as part of production experiment): Correlation matrix for sub-scores

	Verbal density	Lexical diversity	Errorfree
Verbal density	1.000	.596***	.401**
Lexical diversity		1.000	.678***
Errorfree			1.000

** significant at $p < .01$ level
*** significant at $p < .001$ level

Table 17. Proficiency measure (as part of comprehension experiment): Correlation matrix for sub-scores

	Verbal density	Lexical diversity	Errorfree
Verbal density	1.000	.587***	.412***
Lexical diversity		1.000	.667***
Errorfree			1.000

*** significant at $p < .001$ level

For each of the two data sets, three components were needed to account for all the variance in the data. The first component which was extracted accounted for by far the most variance within the original three variables: 70.8% of the variance in the production proficiency measure and 70.6% in the comprehension proficiency measure. The two other components which were extracted had eigenvalues of less than 1.00.[39] This means that they explain less variance than the original variables do. In other words, everything which is common to the original variables is contained within the first component. (See StatSoft (2004) for more details about eigenvalues.)

It is interesting to inspect how much each of the three original variables contribute to the extracted component. This is given in a so-called factor loading for each variable. The higher the factor loading, the more the original variable contributes to the extracted component. The factor loadings for the two components extracted for the production proficiency measure and the comprehension proficiency measure are given in Table 18.

Table 18. Proficiency measure: Factor loadings for original variables

	Component extracted for proficiency measure (as part of production experiment)	Component extracted for proficiency measure (as part of comprehension experiment)
Verbal density	.779	.781
Lexical diversity	.912	.905
Errorfree	.828	.829

It is important to observe that the factor loadings for both measures are almost identical, even though only a small proportion of the subjects is common to both data sets (out of the 45 subjects who took part in the production experiment and the 66 subjects who took part in the comprehension experiment only 12 took part in both). This suggests that the adopted method, that is a principal components analysis of verbal density, lexical diversity and the rate of errorfree utterances, works consistently across different subjects. On the assumption that the sub-scores are valid measures of L2 proficiency,

[39] An eigenvalue denotes the amount of variance in a variable or set of variables which is explained by a component (or factor).

the resulting proficiency scores can be considered to provide an approximate yet relatively robust indication of the subjects' L2 proficiency.

The values for the extracted component were saved as standardised variables and these computed factor values (having a standard normal distribution) are used as the proficiency score for each subject. The individual scores are provided in Appendix B. For the production proficiency measure, they ranged from –2.06 to 2.12 and for the comprehension proficiency measure from –1.95 to 1.70. On the basis of these scores, it will be possible to divide subjects into different proficiency level groups. The details of these groups are provided in the Chapter 5 (§5.4.2.1) and Chapter 6 (§6.3.3.1.1).

4.8 Summary

In the present study, an independent measure of L2 proficiency is required for two purposes. First, it is necessary in order to ensure that any comparison between L2 children and L2 adults is valid, and second, it will be used to derive developmental sequences from cross-sectional data. Given these requirements and the lack of a suitable alternative, a privately developed measure was opted for. Data were collected using a picture description task and following the adopted (narrow) definition of L2 proficiency as 'the ability to produce lexically, morphologically and syntactically complex and accurate utterances in the TL', these data were assessed using measures of complexity and accuracy. Several possible measures were evaluated on the basis of the previous literature and by applying them to original data collected from L1 children and adults using the same (picture description) task.

MLUw was concluded to be an unsuitable measure of L2 morphosyntactic complexity because it continues to develop in older (L1) children and it is highly variable amongst adult native speakers. The alternative measure of verbal density, the average number of finite and non-finite verbs per T-unit, was considered to be more appropriate because although this was also quite variable amongst native-speaker adults, in the child L1 sample analysed here, this measure was not found to differentiate between older L1 children. Lexical complexity was measured using the lexical diversity measure known as Guiraud's index ($V/\sqrt{N}$). On both measures of complexity, L1 children differed from L1 adults. In order to factor out these child ~ adult differences from the child L2 ~ adult L2 comparison, the L2ers' scores on each of these measures were standardised within each (child/adult) group before combining them into the final proficiency score. The complexity measures were combined with a measure of accuracy, namely the rate of errorfree utterances. Detailed criteria were provided for what did and did not count as morphological, syntactic or lexical errors.

The three sub-scores of verbal density, lexical diversity and errorfree utterances were combined using principal components analysis. This statistical method reduces these three variables to one by seeking out the commonalities between them whilst attempting to account for as much of their variance as possible. It resulted in one proficiency (z) score for each subject. Proficiency scores were calculated for two different data sets, corresponding to the production task reported on in Chapter 5 and the comprehension task in Chapter 6. The results of these two analyses were very similar, which suggested that the adopted method is robust.

In the following two chapters, we will see how these proficiency scores are put to use in an analysis of the L2 production and comprehension of scrambled direct objects in Dutch.

CHAPTER 5

THE PRODUCTION OF SCRAMBLED OBJECTS

Introduction

This chapter presents experimental data on the acquisition of direct object scrambling in Dutch by three groups of learners: English-speaking L2 children and L2 adults and L1 children. As outlined in Chapter 1, this thesis attempts to answer the question of whether, broadly speaking, child L2 acquisition, adult L2 acquisition and child L1 acquisition of this particular property of Dutch essentially involve the same process(es). In this chapter, this issue is addressed using an elicited production task to determine, first, whether learners demonstrate knowledge of the interpretive constraints on scrambling and, second, which developmental sequences they pass through in their acquisition of this property of Dutch. The results show that, in accordance with the target grammar, the high and (some) mid proficiency L2 children and L2 adults generally scramble indefinite objects when a specific interpretation is intended, while they leave non-specific indefinite objects in the non-scrambled position. Furthermore, they also consistently scramble definite objects across negation in a context where sentential negation is intended. Using the data from the proficiency measure detailed in the preceding chapter to infer developmental sequences on the basis of these cross-sectional data, it is shown that English-speaking L2 children and L2 adults pass through the same developmental stages in their acquisition of direct object scrambling in Dutch.[1]

The chapter is organised as follows. Section 5.1 briefly recapitulates the previous findings on the L1 and L2 acquisition of scrambling detailed in Chapter 3, as well as restating the predictions to be tested here. The details of the experimental method and procedure, used with all three learner groups, are given in §5.2. Section 5.3 presents the results from the L1 children and § 5.4 those of the L2 children and adults. Finally, in §5.5, the different learner groups are compared with each other.

5.1 Recap: Previous findings and predictions

Previous studies on the production of scrambled objects have observed that learners pass through a "no scrambling" or optional scrambling stage before

[1] Parts of this chapter – and in particular, the discussion of Schaeffer's (2000) account and the predictions it makes for L2 acquisition – have been published as Unsworth (2004a). The data presented there form a subset of the data presented here. The additional data reported on here include more L2 children who have targetlike scrambling and more L2 adults who are in the initial L1 transfer stage (cf. §5.4.2).

they scramble objects in a targetlike fashion. Most of these studies – and the only relevant L2 study – use spontaneous production data, which makes it difficult to determine whether scrambled forms are targetlike in the sense that they are restricted to the relevant types of NPs in the appropriate contexts only. The scarcity of scrambled indefinite objects in spontaneous speech also means that there is relatively little known about the production of this type of object in particular.

Schaeffer (2000) investigated the acquisition of scrambled objects in L1 Dutch using a production task designed to elicit both definite and indefinite NP objects. She found that the youngest children (2-year-olds) scrambled definite NPs across negation significantly less often than the older children (3- to 6-year-olds), that is, there was a leap towards adultlike behaviour between the ages of two and three. The individual data for the 2-year-olds indicated that there were both children who scrambled optionally and children who did not scramble at all. The distinction between the 2-year-olds, on the one hand, and the 3- to 6-year-olds, on the other, was also observed for indefinite NPs. However, as discussed in Chapter 3 (§3.2.1.5), the specific indefinite condition was fundamentally flawed and, consequently, even the adults failed to distinguish between the specific and non-specific indefinite conditions. Hence, Schaeffer's indefinite data must be treated with utmost caution. One of the goals of the present study is to improve upon Schaeffer's original design such that the relevant distinction between the two indefinite conditions is observed.

Recall that Schaeffer (2000) proposes that the acquisition of scrambling depends on the acquisition of a pragmatic principle which states that speaker knowledge and hearer knowledge are always recognised as separate entities. This division is crucial to her approach: Schaeffer claims that distinguishing between speaker knowledge and hearer knowledge allows speakers to establish whether a given NP is (i) non-discourse-related, that is, part of long-term shared knowledge between speaker and hearer or (ii) discourse-related, that is, mentioned in the preceding discourse. This distinction is important because in her analysis, discourse-related and non-discourse-related NPs scramble to different positions, namely Disc(ourse)P and Ref(erentiality)P, respectively. In Chapters 2 and 3, it was argued that Schaeffer's approach faces both empirical and conceptual problems (see §2.3.2.1 and §3.2.1.5 for more details). Nevertheless, it is worth briefly considering the predictions her approach would make for L2 acquisition.

Once L1 children have acquired the relevant pragmatic principle and consequently always distinguish between speaker knowledge and hearer knowledge, scrambling should follow automatically. Given that this principle is language-general, it should be in place in all children at around the same age, irrespective of the language they are acquiring, that is, around three to four years.[2] Given that all of the L2 subjects tested here are older than this age, on Schaeffer's approach, they have the pragmatic prerequisite for scrambling, and consequently it is predicted that, all things being equal, they should not have any problems with scrambling. It is expected that as a result of L1 transfer, however, English-speaking L2ers of Dutch will at first not have the syntactic knowledge necessary for scrambling and, therefore, despite their having the pragmatic principle which Schaeffer claims is necessary for targetlike scrambling, they will not scramble initially. Crucially, however, once they can scramble, they should always do so. This is because, on Schaeffer's account, optional scrambling in L1 children stems from their lack of knowledge of the pragmatic principle which states that speaker knowledge and hearer knowledge are *always* separate, knowledge which, as noted above, the L2ers should have.

In Chapter 3, I argued that for the English-speaking L2ers, the acquisition of direct object scrambling in Dutch constitutes a poverty-of-the-stimulus problem (at least for indefinite NPs). In addition to ascertaining whether any L2ers are able to overcome this problem, the production experiment reported on in this chapter will be used to test predictions ❶ and ❸ formulated in Chapter 3 (§3.5) and repeated below. Recall that these – and in particular prediction ❶ – are based on (i) the assumption that child L2 acquisition is constrained by UG and (ii) the hypothesis that adult L2 acquisition is also constrained by UG.

PREDICTION ❶. CHILD L2 ~ ADULT L2: DEVELOPMENTAL SEQUENCES
L2 children and L2 adults will pass through the same developmental sequence.

PREDICTION ❸. CHILD L2 ~ CHILD L1: DEVELOPMENTAL SEQUENCES
As a result of L1 transfer, L2 children will pass through a different developmental sequence from L1 children.

5.2 Method

The purpose of the production task is to determine whether learners know the interpretive constraints on scrambling by establishing whether they produce the targetlike form (scrambled vs. non-scrambled) in the relevant context (specific vs. non-specific for indefinite objects, sentential vs. contrastive negation for (anaphoric) definite objects). The task is a modified version of Schaeffer's (2000) experiment. As mentioned above, the indefinite conditions in this experiment were problematic and hence they have been modified. Details of this and other modifications are presented in the following section.

[2] Here, *general* means not tied to any particular language.

5.2.1 Materials

The task was a combination of a truth value judgement task and an elicited production task.[3] Subjects were presented with short stories, illustrated in a picture book and with up to three pictures per story. At the end of each story, a silly puppet, *Karel de Kikker* 'Charles the Frog', either made a comment or asked a question. Subjects were asked to help the puppet by saying whether his comment was true or false and if it was false, to correct it, or if he asked a question, they should answer it. The puppet's questions and comments were designed so that in their responses, subjects would produce a scrambled or non-scrambled utterance. There were three conditions: definite NP (target: scrambled), specific indefinite NP (target: scrambled) and non-specific indefinite NP (target: non-scrambled).

The definite condition was more or less the same as Schaeffer's definite condition. The protagonist in the story is going to carry out an action on a particular object, but for some reason, decides not to. The puppet subsequently guesses that the protagonist *is* going to carry out the action in question. This is clearly false and hence subjects are expected to correct the puppet using negation in their response.[4] Note that the stories never contain any (non-)scrambled objects: when objects are used in conjunction with the negator *niet*

---

[3] As will become apparent, the extent to which subjects are actually asked to judge the truth of a particular statement is very limited: this occurs in the definite NP condition only and a subset of the fillers and it never involves subjects judging the truth value of (non-)scrambled sentences. This task is thus ultimately a production task and as such, very different from the (real) truth value judgement tasks which are discussed in Chapter 6.

[4] In Schaeffer's original experiment, the puppet's comment contained a topicalised object and the contrastive modal particle *wel*, as in (i)-a. The inclusion of such a contrastive element was especially important in the conditions which tested scrambling across adverbs. In these, it was expected that providing a contrastive adverb, such as *vlug* 'quickly' in (i)-b, would force subjects to produce an adverb in their responses (here, *langzaam* 'slowly').

(i) a. De appel gaat Pluto wel opeten
the apple goes Pluto up-eat
'Pluto IS going to eat the apple.'

b. Pluto gaat de dinosaurus vlug opeten
Pluto goes the dinosaur quickly up-eat
'The dinosaur is going to eat Pluto quickly.'

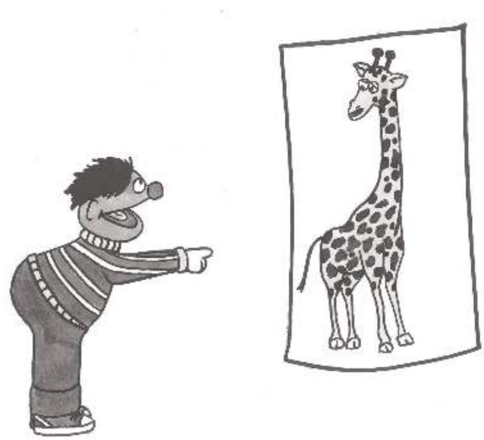
Because the adverb conditions were not particularly successful with Schaeffer's younger (i.e. less proficient) subjects, they are not included here. Given that they are not included and that it is possible to elicit negation by simply providing a false statement (without using *wel*), the puppet's comment in the definite condition was changed to an affirmative sentence with canonical order (see (1) for an example scenario). This was also considered desirable because some of the native speakers who were consulted in the pilot stage of this experiment claimed that the topicalised object sounded infelicitous.

'not', they are topicalised; an example scenario is presented in (1).[5] The expected response is the scrambled utterance in (1)-a. The non-scrambled order in (1)-b is inconsistent with the given scenario because it is interpreted as expressing contrastive negation.

(1) Example scenario for definite NP condition

Context:	**Ernie**	**wil**	**iets**	**natekenen**
	Ernie	wants	something	copy
	'Ernie wants to copy something.'			

[5] Each of the stories involved in this experiment (in all three conditions) deals with an action which is crucially not going to take place. Understanding this is a prerequisite for an informative response (i.e. a response containing all the elements necessary to determine whether scrambling has or has not taken place in a targetlike fashion – the object, the negator and the verb). Conveying the non-occurrence of a particular action is rather tricky, however, especially to younger/low proficiency subjects. In order to make the intended meaning clear, the pictures included thought-bubbles which were struck through with a cross, as the pictures in (1) through (3) illustrate. Thought-bubbles have been employed successfully with young children by other researchers (e.g. De Villiers 2005) and none of the subjects reported on here seemed to have any difficulties understanding the message they were designed to convey.

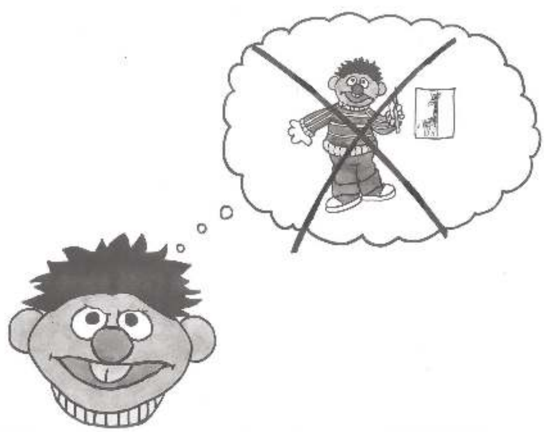


Ernie:	**Kijk,**	**een giraffe.**	**Wat**	**een grote**	**giraffe!**
	look	a giraffe	what	a big	giraffe

'Look, a giraffe. What a big giraffe!'

Hij	**is**	**echt**	**te**	**groot**
he	is	really	too	big

'It's really too big.'



Dus	**die**	**ga**	**ik**	**NIET**	**natekenen**
so	that	go	I	not	copy

'So I'm NOT *going to copy that.'*

Puppet:	**Ernie**	**gaat**	**de giraffe**	**natekenen**
	Ernie	goes	the giraffe	copy

'Ernie is going to copy the giraffe.'

Exp:	**Nee hé?**	**Wat**	**gebeurt**	**er**	**echt?**
	No, right	what	happens	there	really

'No, right? What's really going to happen?'

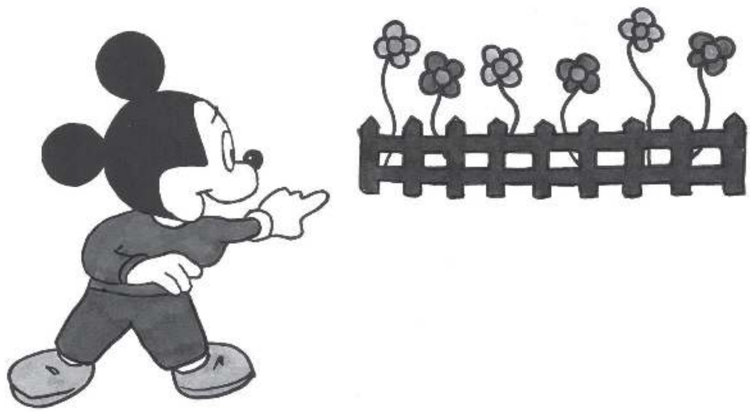
a.	**Ernie**	**gaat**	**de giraffe**	**NIET**	**natekenen**
	Ernie	goes	the giraffe	not	copy
b.	**Ernie**	**gaat**	**NIET**	**de giraffe**	**natekenen**
	Ernie	goes	not	the giraffe	copy

Note that subjects were trained to start their responses with the subject and the auxiliary *gaat* 'goes' to circumvent V2, that is, to ensure that the remainder of the sentence would be informative with respect to scrambling. The younger child subjects were told that this helped the puppet understand better and the older child and adult subjects were simply requested to do so.

The specific indefinite condition involved a story with six objects, three each of two different colours. The protagonist states that s/he would like to manipulate the objects but for some reason (for example, not enough time), s/he decides to leave one unmanipulated. In this condition, the puppet says that he was not paying attention or he pretends not to understand and asks the subject 'What is X (the protagonist) not going to do?' The subject is expected to respond by referring to the one object which is not going to be manipulated. Discussion with native speakers in a pilot stage of the experiment suggested that if the objects are not differentiated in some way, the most natural way to single out one object, which is not going to be manipulated, is by using the cardinal/stressed indefinite *één* 'one' rather than the indefinite determiner *een* 'a'. Given the fact that the English equivalent of *één*, that is 'one', is a possible (albeit not completely unambiguous) means of conveying the specific interpretation of the object NP in the L2 subjects' L1 (cf. §2.2), the use of *één* in the experimental scenario would be undesirable as it might plausibly encourage subjects to use this rather than scramble.[6] By qualifying the object NP with an adjective, the unstressed determiner (pronounced as /ən/) sounds better. In order to make the use of the adjective felicitous, at least two other objects of a different colour must be present. In order to clearly single out one object which will not be manipulated, a group of three is depicted. An example scenario is provided in (2). Note that the non-scrambled response given in (2)-b is false, because it is not true that Mickey will not pick any (red) flowers.

[6] Avoiding the use of *één* in the scenario does not of course preclude subjects using it in their responses and as we shall see, this is borne out in the results (cf. §5.4.3.3).

(2) Example scenario for specific indefinite condition



Context:	**Mickey**	**staat**	**in**	**de tuin**
	Mickey	stands	in	the garden

'Mickey is standing in the garden.'

Mickey:	**Kijk,**	**wat**	**een mooie bloemen,**	**zeg!**
	look	what	a beautiful flowers	say

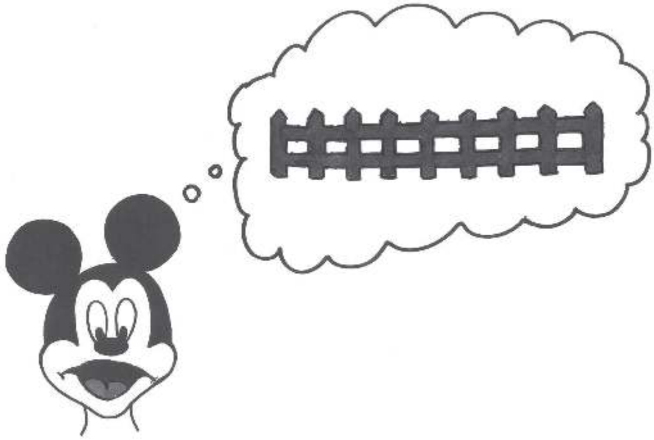
'Look, what beautiful flowers!'

Gele	**bloemen,**	**rode**	**bloemen.**
yellow	flowers	red	flowers

'Yellow flowers, red flowers.'

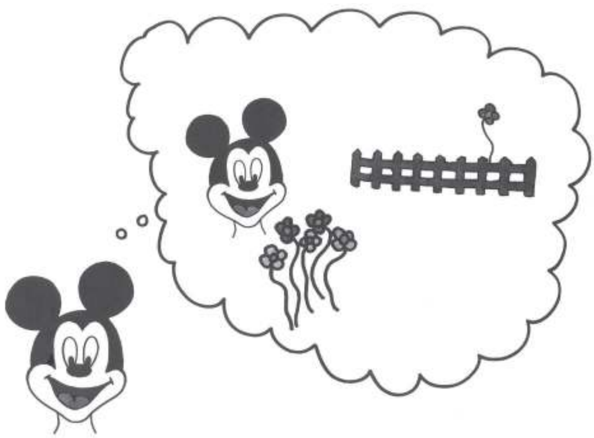
Ik	**hou**	**van**	**bloemen.**	**Ik**	**ga**	**ze**	**plukken**.
I	love	from	flowers	I	go	them	pick

'I love flowers. I'm going to pick them.'



Maar	**dan**	**is**	**de tuin**	**leeg**
but	then	is	the garden	empty

'But then the garden will be empty.'



Dus	**een rode bloem**	**ga**	**ik**	**NIET**	**plukken**
so	a red flower	go	I	not	pick

'So I'm NOT going to pick a red flower.'

Puppet:	**Ik**	**zat**	**niet**	**op**	**te**	**letten**
	I	sat	not	up	to	pay attention

'I wasn't paying attention.'

Puppet:	**WAT**	**gaat**	**Mickey**	**niet**	**doen?**
	What	goes	Mickey	not	do

'WHAT'S Mickey not going to do?'

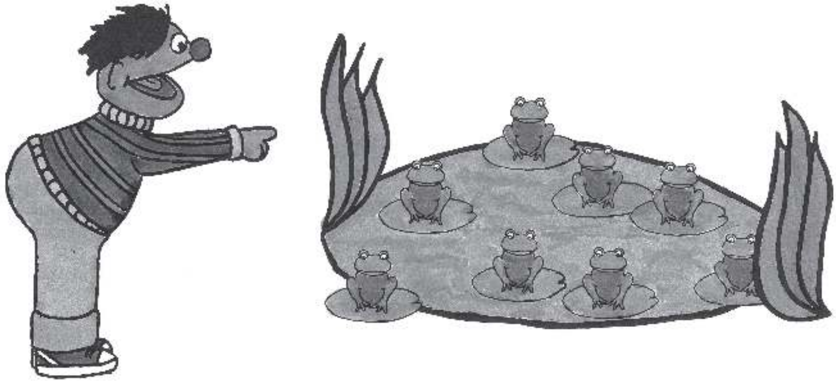
a.	**Mickey**	**gaat**	**een (rode) bloem**	**niet**	**plukken**	
	Mickey	goes	a (red) flower	not	pick	
b.	**Mickey**	**gaat**	**niet een /**	**geen (rode) bloem**	**plukken**	
	Mickey	goes	not a /	no (red) flower	pick	

The set-up of the non-specific indefinite condition is more or less the same as the specific indefinite condition, except that rather than having a small group of objects from which a single object was singled out, there was a large group of items which was designed to discourage subjects from pinpointing one particular object, for example, by making the objects look similar to each other. The protagonist in the story says that s/he is not going to manipulate any of these objects, by stating that s/he is not going to carry out the action in question (*Dat ga ik niet doen* 'I'm not going to do that') without actually mentioning the object. A (non-scrambled) response is elicited with the same question as in the specific condition. An example scenario is provided in (3). The expected response is the non-scrambled order given in (3)-a. Note, however, that the scrambled response, indicating a specific interpretation of the object, is not inconsistent with the pictures used here; rather, it is somewhat unexpected given that there is no particular object which has been singled out in the preceding scenario. In both indefinite conditions, the experimenter sometimes followed up on subjects' responses with the question *Waarom niet?* 'Why not?' in order to check that subjects were paying attention and/or had understood the scenario.

(3) Example scenario for non-specific indefinite condition

Ernie:	**Mmm,**	**ik**	**heb**	**zin**	**om**	**iets**	**te**	**vangen**
		I	have	desire	to	something	to	catch

'Mmm, I fancy catching something.'



Kijk,	**een vijver!**	**Er**	**zitten**	**kikkers**	**in.**
Look	a pond	there	sit	frogs	in

'Look, a pond! There are frogs in it.'

Maar	**een kikker**	**is**	**moeilijk**	**om**	**te**	**vangen**
but	a frog	is	difficult	around	to	catch

'But a frog is difficult to catch.'

Ik	**denk**	**niet**	**dat**	**ik**	**dat**	**kan**
I	think	not	that	I	that	can

'I don't think I can do that.'

Dus	**dat**	**ga**	**ik**	**NIET**	**doen**
so	that	go	I	not	do

'So I'm NOT going to do that.'

Puppet:	**Ik**	**zat**	**niet**	**te**	**luisteren**
	I	sat	not	to	listen

'I wasn't listening.'

Puppet:	**WAT**	**gaat**	**Ernie**	**niet**	**doen?**
	What	goes	Ernie	not	do

'WHAT'S Ernie not going to do?'

a.	**Ernie**	**gaat**	**niet een /**	**geen kikker**	**vangen**
	Ernie	goes	not a /	no frog	catch
b.	**Ernie**	**gaat**	**een kikker**	**niet**	**vangen**
	Ernie	goes	a frog	not	catch

There were six items per condition using the same six verbs: *lezen* 'to read', *plukken* 'to pick', *vangen* 'to catch', *opeten* 'to eat (up)', *uitknippen* 'to cut out' and *natekenen* 'to copy'. These were interspersed with 20 filler items of four different types. Twelve fillers were taken from another experiment on binding. These were included in order to distract subjects (especially the older subjects) from the true purpose of the experiment and to provide a comparable scenario to the definite condition, that is, where a truth value judgement was required. The puppet commented on a picture using a sentence containing the reflexive anaphor *zich* and subjects were asked to state whether he was correct, as in the definite condition. Half of these fillers required a yes-response and half a no-response.[7] Examples are provided in Appendix C.

The remaining eight fillers were designed to check that subjects were paying attention and that they understood the nature of the task. They also required responses which started with the grammatical subject and the auxiliary *gaat* 'goes'. Two fillers had a similar format to the two indefinite conditions; subjects were asked a simple question about a short story. This was the affirmative version of the negative question used in the indefinite conditions, that is, *Wat gaat X doen?* 'What is X going to do?'. A further two fillers were taken from Schaeffer's original experiment. These contained a similar set-up to the definite condition except that they contained an adverbial (*hier* 'here' or *straks* 'later') rather than negation and were designed to elicit a yes-response. The final four fillers had the puppet make a statement about a short story which subjects were required to judge; two were designed to elicit a yes-response and two a no-response. As well as providing another comparable scenario to the definite condition, where the subject is asked to state whether the puppet's statement is true, these additionally served as an extra check that the L2 subjects were of a sufficient proficiency level to complete the task. A complete list of fillers is provided in Appendix C.

5.2.2 Procedure

The experiment was carried out as follows. Subjects were told that they were going to listen to stories about some pictures in a book (a ring-binder containing colour pictures inserted into see-through plastic sleeves). They were then introduced to the silly puppet, *Karel de Kikker* 'Charles the Frog'. The puppet played a very active role with the younger children, but for the older children and adults, he was used to explain what the subjects should do and subsequently placed to one side. Specifically, the subjects were shown how the puppet was used with the younger children and once this was clear, the experimenter said that she would play the role of the puppet. The experiment started with a short warm-up session to familiarise subjects with the task. Subjects were presented with simple stories using pictures; these used a comparable set-up to many of the fillers. In order to proceed to the main task,

---

[7] After the first round of testing was complete, it became clear that the pictures used in these fillers were slightly problematic (see Coopmans, Krul, Planting, Vlasveld and van Zoelen 2004 for details). Only the L1 children and those L2 children who participated in both the production and the comprehension tasks (see Chapter 7) saw the modified (i.e. corrected) version of the pictures. As noted above, these fillers were included to distract the subjects from the real goal of the experiment and to provide a comparable scenario to the definite condition, and hence the fact that they were slightly problematic does not matter for present purposes.

they had to provide targetlike responses to a minimum of three items, that is, they had to provide at least one correct no-response, one correct yes-response and correctly answer one question. As many items as necessary (up to a maximum of six) were used until the subject felt comfortable with and understood the nature of the task. When in the warm-up session subjects failed to provide a targetlike response, the experimenter helped them and explained the purpose of the task (or 'the rules of the game' for the child subjects) once again. In particular, it was necessary to ensure that participants understood that they should start their responses with the subject and the auxiliary *gaat* 'goes'. Once this was established, the experimenter proceeded to the main experiment. The pictures for the warm-up session and the experiment were in the same book to ensure that it was possible to proceed seamlessly from one part to the other. The main experiment was presented in two pseudo-random presentation orders, distributed as evenly as practicably possible across subjects.[8, 9] Each subject was tested individually in a quiet room (at the subject's school or at Utrecht University or the University of Amsterdam) and each session was recorded using a Sony Hi-8 video recorder and a Tascam DA-P1 DAT recorder with a Crown PZM185 microphone. Child L1 subjects were rewarded for their participation with a small snack, the child L2 subjects with a small gift (such as stickers or a pen) and the adult subjects with a piece of cake.

With the details of the method and procedure in place, we now turn to the results. For expository reasons, we start with the L1 children, although the L2 data, which were also collected first, will form the bulk of the chapter.

5.3 L1 acquisition

The purpose of carrying out the production task with L1 children is twofold. First, given the problems with the indefinite conditions in Schaeffer's original experiment, new L1 data are needed to truly establish whether L1 Dutch children know the interpretive constraints on scrambled indefinite NP objects. Second, new L1 data from a task employing exactly the same experimental method as with the L2 subjects are necessary to maximise the comparability of the data used in the child L2~child L1 comparison.

---

8 There was only one set of pictures and the order needed to be changed by hand. Time constraints meant that it was sometimes impossible to change the order between subjects and hence the number of subjects given each order is slightly unevenly distributed across subjects.

9 The items were divided into two parts. Order B consisted of the reverse order of order A in part I, followed by the reverse of order A in part II. Constructing the different presentation orders in such a way meant that if, because of time constraints or fatigue, subjects could not complete the task, subjects would always be tested on the same items and hence comparability across subjects was ensured.

5.3.1 Subjects

Data were collected from a group of 5-year-old monolingual Dutch children (n=13; range = 5;6,2 – 5;11,30; mean = 5;8). This age group was selected for two reasons. First, the indefinite data in the younger children in Schaeffer's experiment were very limited. The youngest children in her study produced relatively few (and sometimes no) indefinite NP objects. Therefore, slightly older children were selected here because they might be expected to produce more indefinites and scramble consistently. Furthermore, we know from previous research that at age 5, children interpret scrambled indefinite objects in a non-targetlike fashion (cf. §3.3.1). Thus, by collecting production data from this group, it will additionally be possible to establish whether there is a discrepancy between production and comprehension, as earlier work suggests.[10] The task was also carried out with native-speaker adult controls

---

[10] Data were also collected from a group of 4-year-olds (n=15) and another group of 5-year-olds (n=7), but these were unfortunately discovered to be unusable due to a recurring experimenter error. This error was detected as a result of the unexpectedly high percentage of *non*-scrambled definite objects (76.6% (49/64) across both groups), many of which contained the sequence *geen* + N (which is expected with a non-scrambled non-specific indefinite NP; cf. (3)). The problem was that throughout the experiment – probably as a result of unclear instructions on my part – the experimenter responsible for eliciting the responses from the children consistently started children's responses with the subject and the auxiliary *gaat* 'goes'. This appears to have altered the focus structure of the elicited sentence in the following way. By starting the sentence given in (1) for the child with *Ernie gaat*, the experimenter is implicitly telling the child: 'Say something about Ernie, but it doesn't necessarily have to be about the giraffe.' In other words, by producing *Ernie gaat*, the experimenter indicates to the child that Ernie is old information, implying that the child should provide the new information, that is, what it is that Ernie is not going to do. If the definite NP object, *de giraffe*, were also to be considered old information, this would be provided, too (but it is not, of course, because this would reveal the target response). If *de giraffe* is not considered part of the old information, the fact that it is a definite NP arguably becomes irrelevant and it can thus either be left unscrambled as a definite NP or be changed to an indefinite with *geen*. In this way, the child is saying that 'what Ernie isn't going to do is any copying of giraffes' (rather than 'as for that giraffe, Ernie's not going to copy it', which is the expected reading when the object is considered old information). I have presented several native speakers with the experimental scenario above and they concur that if the response were started for them, they would (or at least could) use a non-scrambled or *geen* form. Of course such a response is in principle not impossible when the subject starts her own sentence. The point is that when the subject's utterance is started by the experimenter, a non-scrambled- or *geen*-response is much more likely. This is confirmed by the following observation: out of the 23 responses in the definite condition containing a non-scrambled definite NP or *geen* form, 20 (87.0%) were started by the experimenter. These responses are from children who produce at least one utterance containing a scrambled object in this or the specific indefinite condition, which indicates that these non-scrambled forms do not result from a lack of knowledge of scrambling. The experimenters who tested the subjects reported on in the main text were given explicit training on how to avoid this problem. The fact that the number of non-scrambled objects in the definite condition was consequently significantly reduced, from 68% (17/25; for the 5-year-olds only) to 28.2% (20/71), suggests that this experimenter error was indeed the source of the unexpected results discussed here.

(n = 11; range = 18-24; mean = 20;6), most of whom were students (but not in linguistics) at Utrecht University.

The experiment was carried out by two native-speaker student assistants with the L1 children and by one native-speaker student assistant and the present author (near-native speaker) with the L1 adults. One experimenter told the stories and elicited the responses from subjects and the other played the role of the puppet. The whole experiment lasted between 20 and 30 minutes; for the children, it was split into two shorter parts of approximately 10-15 minutes.

5.3.2 Results

The data will be analysed in terms of the percentage of objects scrambled in each of the three conditions. The results are first presented for the two groups (5-year-olds and adults) as a whole in §5.3.2.1 and subsequently for the individual children in §5.3.2.2. First, we briefly consider the fillers. Two fillers consistently elicited unexpected responses, from children and adults alike. In these two items, both taken from Schaeffer's original experiment (cf. §5.2.1), subjects were expected to accept a sentence, a true description of the story in question, with a topicalised object. An example is given in (4).

(4)	De olifant	gaat	Nijntje	straks	natekenen
	the elephant	goes	Miffy	later	copy
	'Miffy is going to copy the elephant later.'				

On numerous occasions, however, both children and adults interpreted the topicalised object as the subject (i.e. that the elephant is going to draw Miffy) and hence rejected the statement, saying something like 'Miffy is going to copy the elephant later (and not the other way round)'. There are two factors which may have contributed to this unexpected response: first, native-speaker informants tell me that the intended interpretation is most easily accessed with a very specific intonation (with a slight pause after the object) and this may have been absent in some cases; second, in this particular example, the topicalised object is animate and hence could easily be interpreted as the agent. Given that native-speaker adults both rejected and accepted these fillers, both responses were counted as targetlike for all other subjects. Out of the remaining fillers (excluding those on binding – cf. fn. 7), there was just one which was answered incorrectly (E2 by L1#7P).[11]

There was no effect of presentation order in any condition (t-test (2-tailed; equal variances not assumed): definite condition, $t = -.029$, $p = .978$; specific indefinite condition, $t = .237$, $p = .818$; non-specific indefinite condition, $t = -1.228$, $p = .280$).

[11] The code L1#7P denotes that this is L1 child number 7 in the production (P) experiment.

5.3.2.1 Group results

The group results for both children and adults are provided in Table 19.[12]

Table 19. L1 children and L1 adults (production): Group results. Average percentage scrambled objects

Group	n	Condition					
		Definite (target: scrambled)		Specific indefinite (target: scrambled)		Non-specific indefinite (target: non-scrambled)	
		%	SD	%	SD	%	SD
5-year-olds	13	71.7	30.7	61.3	35.0	15.2	31.3
Adults	11	98.5	5.0	92.9	12.6	0.0	0.0

Let us first consider the adult controls. They consistently scramble in the definite condition and they make a clear distinction between the specific and the non-specific indefinite conditions, scrambling almost always in the former and never in the latter. The difference between these two conditions is statistically significant (Wilcoxon signed ranks test: Z = -3.064, p = .002), as is the difference between the non-specific indefinite and definite conditions (Wilcoxon signed ranks test: Z = -3.027, p = .001).[13] The difference between the specific indefinite and the definite conditions is not significant (Wilcoxon signed ranks test: Z = -1.461, p = .144).

As for the L1 children, their responses show the same (relative) pattern as the adults. Definites are scrambled most of the time, although not quite as frequently as in the adult data, and there is a clear distinction between the specific indefinite condition and the non-specific indefinite condition. The difference between the two indefinite conditions is statistically significant (Wilcoxon signed ranks test: Z = -2.677, p = .007). The specific indefinite and definite conditions do not differ from each other (Wilcoxon signed ranks test: Z = -.489, p = .624) but the definite and the non-specific indefinite conditions do (Wilcoxon signed ranks test: Z = -2.812, p = .005).

Comparing the child and the adult groups, the two groups do not differ significantly from each other in the non-specific indefinite condition (Mann-

---

[12] The figures are presented as the *average percentage* of scrambled responses per group because, as a result of the number of tokens differing per subject, it is this figure which is used in the statistical tests. The more conventional calculation of a straight percentage of scrambled objects per group (i.e. total number of scrambled objects per group divided by total number of scrambled and non-scrambled objects per group) gives similar results: for the 5-year-olds, the rate of scrambling is 71.8% (51/71) in the definite condition, 60.7% (34/56) in the specific indefinite condition and 12.5% (3/24) in the non-specific condition, and for the adults, this is 98.3% (59/60) in the definite condition, 95.5% (42/44) in the specific indefinite condition and 0% (0/41) in the non-specific condition.

[13] The non-parametric Wilcoxon signed ranks test is used to determine whether two paired, non-normally distributed or small (< 30 cases) samples stem from the same population.

Whitney test (2-tailed): Z = -1.734, p = .083), although the p-value is approaching significance (probably because the adults never scramble in this condition, whereas there are three scrambled tokens in the child data). There are statistically significant differences between the two groups in the specific indefinite condition and in the definite condition (Mann-Whitney test (2-tailed): Z = -2.417, p = .016 and Z = -2.702, p = .007).[14] Despite being significantly different from the adults in these two conditions, the children are clearly on their way to adultlike levels of scrambling. The children scramble at a rate of 71.7% in the definite condition and at a rate of 61.3% in the specific indefinite condition. These figures could potentially result from an averaging effect, that is, there could be two or three different sub-groups of children with different response patterns. Hence, it is necessary to examine children's individual results. Furthermore, an analysis of individual data will also form part of the child L2~child L1 comparison.

5.3.2.2 Individual results

The individual results for the adults are given in Appendix C and for the L1 children in Table 20 below. Despite each condition containing the same number of items, the number of tokens for each condition in Table 19 is different. This is because in the two indefinite conditions, subjects sometimes used definite NPs rather than indefinites and a number of alternative answers which did not involve scrambling also occurred. Examples for the specific indefinite condition and the non-specific indefinite condition are given in (5) and (6), respectively. When subjects responded in this way, the experimenter playing the puppet pretended not to understand so that they would reformulate their answer, in the hope that this would lead to a useable response. This was a relatively successful tactic.

(5) Alternative responses in the specific indefinite condition

a. Mickey gaat dat ene bloemetje niet plukken
Mickey goes that one flower not pick
'Mickey isn't going to pick that one flower.'

b. Mickey gaat niet alle bloemetjes plukken
Mickey goes not all flowers pick
'Mickey isn't going to pick all flowers.'

c. Mickey gaat er ééntje laten liggen
Mickey goes one leave lie
'Mickey is going to leave one of them."

d. Mickey gaat het laatste bloemetje niet plukken
Mickey goes the last flower not pick
'Mickey isn't going to pick the last flower.'

---

[14] The non-parametric Mann-Whitney test is used for data which do not conform to the prerequisites for a t-test.

(6) Alternative response in the non-specific indefinite condition

a. Ernie gaat de kikkers niet vangen
Ernie goes the frogs not catch
'Ernie isn't going to catch the frogs.'

b. Ernie gaat al die kikkers niet vangen
Ernie goes all those frogs not catch
'Ernie isn't going to catch all those frogs.'

The L1 adults also produced such responses in these two conditions. The 'other' column in Table 20 gives these alternative responses. One token was excluded for subjects L1#6P and L1#12P in the specific indefinite condition and for subjects L1#2P, L1#3P, L1#4P, L1#9P and L1#12P in the non-specific indefinite condition. This was either because the subject misunderstood the scenario or because of factors beyond the experimenter's control, such as an interruption in the test procedure.

In the definite condition, ten out of the 11 adults always scramble and the remaining subject scrambles in all but one item (83.3% (5/6)). Nine of the 11 adults always scramble in the specific indefinite condition; the remaining two subjects scramble in all but one item (75% (3/4) each). No subjects scramble in the non-specific indefinite condition.

Table 20. L1 children (production). Individual results: Percentage of objects scrambled

Subject	**Age**	**Definite NP condition**			**Specific indefinite condition**				**Non-specific indefinite condition**			
		%	**n**	**Other**	**%**	**n**	**Other**	***geen*-forms**	**%**	**n**	**Other**	***geen*-forms**
L1#1P	5;11.10	75	3/4	1	75	3/4	2	1/4	0	0/3	3	3/3
L1#2P	5;11.14	100	6/6	0	16.7	1/6	0	5/6	0	0/4	1	4/4
L1#3P	5;6.02	100	6/6	0	0	0/6	0	4/6	0	0/3	2	3/3
L1#4P	5;7.28	83.3	5/6	0	100	4/4	2	0/4	100	1/1	4	0/1
L1#5P	5;6.20	33.3	2/6	0	25	1/4	2	0/4	---		6	---
L1#6P	5;10.02	40	2/5	1	100	4/4	0	0/4	50	1/2	4	1/2
L1#7P	5;10.01	83.3	5/6	0	83.3	5/6	0	0/5	0	0/1	5	0/1
L1#8P	5;9.04	100	6/6	0	50	1/2	4	0/2	0	0/1	5	1/1
L1#9P	5;11.30	33.3	2/6	0	50	2/4	2	2/4	0	0/3	2	2/3
L1#10P	5;8.14	100	6/6	0	100	6/6	0	0/6	33.3	1/3	3	2/3
L1#11P	5;8.19	100	5/5	1	66.7	2/3	3	0/3	0	0/1	5	1/1
L1#12P	5;7.21	66.7	2/3	3	100	4/4	2	0/4	0	0/1	4	1/1
L1#13P	5;9.01	16.7	1/6	0	33.3	1/3	3	0/3	0	0/1	5	0/1

As for the children, in the definite condition, five of the 13 scramble in all items and four in all but one of the items. The rate of scrambling for the remaining four children is between 16.7% (1/6) and 40% (2/5) items. The number of tokens per child is generally much lower in the two indefinite conditions than in the definite condition, especially in the non-specific condition. This is due to the use of (acceptable) alternative responses (coded as 'other') exemplified above. In the specific indefinite condition, four children scramble in all items and three do in all but one of the items.[15] Two of the 13 children never scramble, and the rate of scrambling for the remaining four children ranges from 16.7% (1/6) to 50% (2/4) of items.[16]

In the non-specific condition, most of the children (10/13) never scramble. The remaining three children produce one scrambled token each. For one subject, this was the only token in this condition, and the other two subjects further produced, respectively, one and two non-scrambled tokens. It should be noted that in this condition, there were six subjects who produced one token only. Across all children, the non-scrambled utterances were generally (75% (18/24)) realised as *geen* ((7)-a), as opposed to *niet een* ((7)-b). This is given for each child in the *geen*-forms column. (See §2.1.2 for why the *geen* form is analysed as non-scrambled.) There is one child, L1#3P, who produces the *geen* form and the *niet een* form. The *geen* form was also used in the specific indefinite condition, accounting for just over half (54.5% (12/22)) of the non-scrambled indefinites. These were produced by four subjects.

(7)	a.	Ernie	gaat	geen	kikker	vangen
		Ernie	goes	no	frog	catch
	b.	Ernie	gaat	niet	een kikker	vangen
		Ernie	goes	not	a frog	catch

It has been suggested (Krämer 2000:178, in response to Schaeffer's data) that the L1 children who fail to scramble in the specific indefinite condition might actually intend a non-specific interpretation. Admittedly, in the course of the experiment, it is often difficult to be sure that subjects understand the context as it was intended. In theory, when subjects produce a non-targetlike response, the experimenter might probe them in order to determine whether any non-targetlike productions, such as those in (7), actually have the meaning conveyed in the context. In doing so, however, the subject is likely to wonder why/if they have said something 'wrong', and possibly as a consequence, s/he might become unsure of his or her (subsequent) response(s). This tactic was

---

15 One of these subjects (L1#8P) scores 50% (1/2).

16 One subject (L1#7P) produced one utterance with double negation, as in (i). This item was excluded.

(i)	Mickey	gaat	niet	een bloem	niet	plukken
	Mickey	goes	not	a flower	not	pick
	'Mickey isn't going not to pick a flower.'					

therefore not pursued. There are, however, a few responses produced spontaneously by subjects who also failed to scramble, which indicate that they understand the specific indefinite scenarios as intended. These are given in (8). The response in (8)-b, for example, involves a non-scrambled demonstrative object with contrastive stress on the determiner. This utterance means that Bert is not going to catch one particular fish and at the same time presupposes that there are fish which will be caught, thereby demonstrating the child's targetlike understanding of the scenario. The responses in (8) do not of course provide direct evidence that the subjects in question also understood in a targetlike fashion those scenarios in response to which they produced non-scrambled utterances, but given the similarity in set-up across the different items in this condition, it seems plausible to assume that this targetlike understanding can be extended to other items in the same condition.

(8) a. Donald gaat die ene olifant niet natekenen (L1P#5)
Donald goes that one elephant not copy
'Donald isn't going to copy that one elephant.'

b. Bert gaat niet DIE vis vangen (L1P#8)
Bert goes not THAT fish catch
'Bert isn't going to catch THAT fish.'

c. Winnie de Poeh gaat dat ene cakeje niet opeten (L1P#9)
Winnie the Pooh goes that one cake not up-eat
'Winnie the Pooh isn't going to eat that one cake.'

A comparison of the individual results in the two indefinite conditions shows that virtually all (10/11) children who produce scrambled objects in the specific indefinite condition and who also produce tokens in the non-specific condition make the targetlike distinction, scrambling the object (more often) in the specific indefinite condition and leaving it unscrambled in the non-specific indefinite condition. For most children (9/11), this distinction is categorical and for two, it is relative. The exception is one child (L1#4P) who scrambles all the time in both conditions; he only produces one token in the non-specific indefinite condition, however.

When the specific indefinite and definite conditions are compared, the following patterns are observed. All but one of the children (12/13) scramble in both conditions some or all of the time. The remaining subject (L1#3P) always scrambles (6/6) in the definite condition but never (0/6) in the specific indefinite condition. There is no discernible pattern in the rate of scrambling in these two conditions relative to each other: there are subjects who scramble relatively more often in the definite than in the specific indefinite condition (L1#2P, L1#8P and L1#11P), subjects who scramble relatively more often in the specific indefinite than in the definite condition (L1#4P, L1#6P and L1#12P), subjects who scramble more or less equally often in both conditions (L1#1P, L1#5P, L1#7P, L1#9P, L1#13P), albeit to differing degrees, and one subject (L1#10P) who scrambles all the time in both conditions. Data from younger children might help establish whether, for example, definite NP objects are scrambled before indefinite NP objects.

5.3.3 Discussion

The native-speaker adult controls pattern as expected. They consistently scramble in the definite and specific indefinite conditions but never in the non-specific indefinite condition. This categorical difference between the two indefinite conditions shows that the modifications applied to Schaeffer's original experiment were successful. The relative differences between the three conditions is reproduced in the child L1 data, both as a group and, on the whole, across individuals. In other words, the 5-year-old children know which objects should be scrambled (definites and specific indefinites) and which should not (non-specific indefinites), that is to say, they are aware of the interpretive constraints on scrambling.

There are two possible objections to this conclusion regarding the L1 children. First, there are some (albeit only three) scrambled objects in the non-specific indefinite condition. In this reagard, it should be pointed out that the production of a scrambled indefinite object in the non-specific condition might be unexpected but it is certainly not ungrammatical. In other words, the three scrambled non-specific indefinites should not necessarily be viewed as errors. As noted above (§5.2.1), the pictures in this scenario were designed to discourage subjects from singling out a particular object, but this of course does not preclude subjects from doing so. It is possible, therefore, that the three tokens (out of a total of 24) which were scrambled in this condition were intended to have the targetlike specific meaning. Note, furthermore, that the three children responsible for these tokens also scramble (some or all the time) in the specific indefinite and definite conditions. In other words, it is not the case that these children randomly produce scrambled objects in this condition but in no other condition. A second possible objection is that there is only one child who *always* scrambles in both the definite and the specific indefinite conditions. We might expect more children to scramble all the time in the necessary conditions, if they really knew the interpretive constraints on scrambling. While more data are certainly needed to establish the age at which all L1 children scramble consistently, the crucial observation made here is that although the majority of children might not scramble all the time, when they do scramble, it is in the conditions where they are supposed to (with the exception of the three tokens discussed above). This suggests that although they still have to attain targetlike levels of scrambling, they are aware of the interpretive effects instantiated in the different word orders. In other words, they use the targetlike form (scrambled vs. non-scrambled) in the appropriate context (specific vs. non-specific interpretation of the indefinite object, sentential negation interpretation for the (anaphoric) definite object).

The L1 data collected here are completely consistent with those reported in Schaeffer (cf. §3.2.1.5). Schaeffer's 5-year-olds (n=10) scrambled definite NPs at a rate of 76.5% (26/34) of items, compared with 71.8% (51/71) for the children (n=13) here. An examination of the individual data collected here indicates that that there are children who scramble consistently (i.e. all the time), children who scramble optionally, and children who never (or hardly ever) scramble, both in the definite and the specific-indefinite condition. This suggests that children pass through a "no scrambling" stage (*contra* Schaeffer 2000; see in particular the discussion of Schaeffer's individual data in §3.2.1.5), before scrambling optionally and, subsequently, all the time. Combining Schaeffer's data (for definite NPs) with the data collected here suggests the developmental sequence given in (9). More cross-sectional data and longitudinal data using the modified experiment are necessary to confirm this.

(9)	**Stage**	**Description**	**Word order patterns produced**
	I	No scrambling	Negation-Object-Verb only
	II-a	Optional scrambling	Object-Negation-Verb *and* Negation-Object-Verb
	II-b	Scrambling	Object-Negation-Verb only

In the non-specific indefinite condition, the L1 children tested here virtually always used the same *geen N*-form. Schaeffer's data are unreliable in this condition, but the data from the 2-year-old child in Hoekstra and Jordens (1994) discussed in §3.2.1.2 suggest that prior to acquiring the *geen*-form, the non-scrambled form with *niet een* is used. This leads to the following hypothesised developmental sequence for the L1 acquisition of (the non-scrambling of) non-specific indefinite NPs:

(10)	**Stage**	**Description**	**Word order patterns produced**
	I	Targetlike non-scrambled order	Negation-*een N*-Verb only
	II-a	Optional use of target lexical form	*geen N*-Verb *and* Negation-*een N*-Verb
	II-b	Targetlike non-scrambled order and form	*geen N*-Verb only

To summarise: The modified elicited production experiment was successful: L1 adult controls consistently scrambled in the definite and specific indefinite conditions and never scrambled in the non-specific indefinite condition. The 5-year-old L1 children tested here make the same (relative) distinctions as the adults, on both a group and an individual level, indicating that they are aware of the interpretive constraints on scrambled objects. The group data suggest that definite NP objects are scrambled before indefinite NP objects, but this is not confirmed by the individual data. By combining the data which were collected here with data available from previous studies, a developmental sequence for the L1 acquisition of scrambling in Dutch was proposed.

5.4 L2 acquisition

Exactly the same production experiment was used to collect data from English-speaking children and adults acquiring Dutch as L2. The core of the present chapter, these data will also subsequently be compared with the child L1 data presented in the previous section.

The data collection procedure for the L2 subjects started with a brief explanation of the general purpose of the experiment, namely comparing younger and older second language learners. For the L2 children, who were usually excused from their regular lessons to take part, this helped to put the subjects at ease. The fact that subjects could speak two languages was always commented on positively in order to emphasise the positive nature of the experiment. They were also told that it was not a test and that what they did had no repercussions for their schoolwork. Subjects were subsequently asked a few general questions regarding their background and language learning history. This allowed us to confirm that subjects fulfilled the criteria for participation (raised in a monolingual English environment, age of arrival after four years, etc.). Nine children were excluded because they had been raised speaking another language in addition to English, and there were five children whose Dutch proficiency level was too low to complete the task. Another five other children were excluded because they had been exposed to Dutch before the age of four.

Before the scrambling task, the L2 subjects completed the picture description task used to measure proficiency (see Chapter 4 for details). In total, each session lasted between 30 and 45 minutes, depending on the subjects' proficiency and age. For the younger children, the scrambling task was conducted over two shorter sessions. In all other respects, the procedure was as laid out in §5.2.2.

5.4.1 Subjects

Subjects were recruited from three different international schools in The Netherlands (where the language of instruction is English) and via personal contacts as well as through advertising on a mailing list for English-speakers in The Netherlands. All the child subjects and all but two of the adult subjects (A8P and A18P) had received some Dutch language instruction, either at the international schools they attended or in foreign language courses and they had varying amounts of contact with Dutch. The children received between 40 and 120 minutes of Dutch language instruction per week at school. All of the subjects had been raised in monolingual English-speaking families. Some of the adults had learned one or more foreign languages at school, but none had started before the age of 12, and for those subjects who were (relatively) fluent in another language, Dutch was still the L2 in which they were the most proficient.[17] All (child and adult) subjects regularly used English. Individual biodata, including knowledge of other languages and amount of contact with Dutch, are provided in Appendix C.

There were 25 L2 children. As per the definition of child L2 acquisition given in Chapter 1, their age at first exposure to Dutch was between 4;0 and 7;1 (mean = 5;6; SD = 1;0). Their age at the time of testing ranged from 5;3 to 17;4 (mean = 9;3; SD = 2;4) and their length of exposure from 0;2 to 13;0 (mean = 3;8; SD = 2;6). Most of the children came from middle class families who had moved to The Netherlands for professional reasons. There were 23 L2 adults. Their age of first exposure to Dutch was between 8;0 and 32;0 (mean = 19;3; SD = 8;8) and their age at the time of testing between ten and 50 years (mean = 23;10; SD = 11;0). They had been resident in The Netherlands for between three months and 27 years (mean = 4;4; SD = 6;0).[18] The vast majority had attended higher education.

5.4.2 Results

The results are analysed as follows. The child L2 and adult L2 groups are divided into three different proficiency levels. In §5.4.2.1, the average rate of targetlike responses per group is calculated for each of the three conditions and the results for the different conditions are compared. An analysis of the individual response patterns is presented in §5.4.2.2. Prior to this, the results for the fillers are discussed.

There were just three fillers in total which were answered incorrectly, by three different subjects (A3P, A10P and A14P).[19] There was no effect of presentation order in any of the three conditions (definite condition: t-test (equal variances not assumed) $t = .256$, $df = 57$, $p = .799$; non-specific indefinite condition: t-test (equal variances not assumed) $t = 1.183$, $df = 42$, $p = .246$; specific indefinite condition: t-test (equal variances assumed) $t = -.673$, $df = 52$, $p = .504$).[20]

---

[17] Ideally, one would like to have subjects who speak no other languages apart from English and Dutch, but given that Dutch is very rarely (if ever) taught as a foreign language in high schools in English-speaking countries and that many adults learn a foreign language at school, this is nigh on impossible. This intervening variable will be examined in §5.4.2.2.4.

[18] Recall that one of the goals of this study is to determine the developmental sequences through which L2 children and adults pass. In order to do this, it is necessary to test subjects who are at different proficiency levels in Dutch, and given that, usually, proficiency is broadly correlated with length of exposure, this means testing subjects with different lengths of exposure.

[19] For reasons outlined above (§5.2.1 and §5.3.2), non-targetlike responses to the fillers on binding and the fillers taken from Schaeffer's original experiment were not taken into account.

[20] This also holds for each proficiency group separately.

5.4.2.1 Results per proficiency group

The child L2 and adult L2 groups were divided into three proficiency groups – low, mid and high – on the basis of proficiency scores, calculated using the protocol detailed in Chapter 4. Using the distribution of scores as a guide to obtain relatively even groups, the subjects were divided as follows: the low proficiency subjects had a score below –0.50, the high proficiency group scored above 0.50 and the mid group's scores fell in-between these two figures.[21, 22] An overview of the proficiency groups is given in Table 21.

Table 21. Overview of proficiency groups (for production experiment)

	L2 children			L2 adults			Mann-Whitney tests comparing child/adult groups
	n	mean	SD	n	mean	SD	
Low[23]	12	-1.06	47	9	-0.92	34	Z = -.444, p = .657
Mid	6	0.10	34	6	0.08	28	Z = -.160, p = .873
High	7	1.35	55	8	0.94	23	Z = -1.390, p = .165

Mann-Whitney tests (two-tailed) were carried out to check that the child L2 group and the adult L2 group at each proficiency level do not significantly differ from each other. The results are given in the final column of Table 21. There is no significant difference between the child and adult groups at any of the three levels. The relative distribution of the three proficiency levels within the L2 child group does not differ significantly from that of the L2 adult group (Chi-squared: χ^2 = .413, df = 2, p = .814).

5.4.2.1.1 Definite condition

Table 22 presents the average scrambling rate for each proficiency level in the L2 child and L2 adult groups. Recall that the targetlike response in this condition is scrambled.

---

21 The proficiency scores are z-scores. This means that scores for most subjects (taking children and adults together) will fall around 0. For this reason, the range of scores for the mid proficiency group, which includes 0, is smaller than for the other two proficiency groups. This allows us to differentiate between the subjects adequately.

22 The labels 'low', 'mid' and 'high' are intended as a means of distinguishing the relative L2 proficiency of this sample of subjects only and as a practical nomenclature only. It is not clear to what extent, for example, the low group would generally be considered as a beginner's group and the mid group as intermediate, etc. Crucially, however, the high group is more proficient than the mid group and the mid more proficient than the low group.

23 This includes two L2 children and one L2 adult whose proficiency in Dutch was so low, they were unable to carry out the picture description task. These subjects do not have a proficiency score but they are included in the low proficiency group in this and subsequent tables.

Table 22. L2 children and L2 adults (production): Average percentage targetlike (scrambled) objects in definite NP condition (organised by proficiency group)

	L2 children			L2 adults			Mann-Whitney tests comparing child/adult groups
	n	%	SD	n	%	SD	
Low	12	22.1	34	9	19.3	33	Z = -.039, p = .969
Mid	6	80.0	40	6	88.9	27	Z = -.631, p = .528
High	7	88.1	31	8	87.5	35	Z = .000, p = 1.000

Both the mid and high proficiency children and adults consistently scramble the definite object NP (cf. adult natives who scramble on average at a rate of 98.5%; see §5.3.2). This is in sharp contrast to the low proficiency groups who on average only scrambled in approximately one fifth of items. When subjects failed to scramble, they produced one of two types of utterance, illustrated in (11)-a and (11)-b.

(11) a. Ernie gaat niet de giraffe natekenen
Ernie goes not the giraffe copy

b. Ernie gaat niet natekenen de giraffe
Ernie goes not copy the giraffe

In the sentence in (11)-a the direct object *de giraffe* 'the giraffe' appears directly to the left of the main verb, *natekenen* 'to copy'. In the sentence in (11)-b, the main verb directly follows the negator, that is, the sentence constituents relevant to scrambling appear in a Negation-Verb-Object order.

Some subjects (in all levels) turned the definite NP into a pronoun. In the 20 items where this occurred, the pronominal object was always scrambled, as in (12). A quarter of these utterances were produced by child L2 subjects and the remainder by L2 adults.[24]

(12) Ernie gaat hem niet natekenen
Ernie goes him not copy
'Ernie is not going to copy it.'

[24] Two subjects (C24P and A13P, where C denotes a child L2 subject, A an adult L2 subject and P that this is the production experiment) also produced utterances (three each) with *geen*, as in:

(i) Ernie gaat geen giraffe natekenen
Ernie goes no giraffe copy
'Ernie is not going to copy any giraffe.'

One such response was produced by an L1 child (L1#6P) and one by an L1 adult (N#6P) (see also fn. 10 on the use of this form in the definite condition). When subjects C24P and A13P produced a definite object NP in their remaining utterances, it was always scrambled.

The standard deviations reported in Table 22 suggest that there is some overlap between the low and the mid/high groups and that there is variation within each group. This variation, which seems to be at comparable levels across child and adult groups, is discussed below (§5.4.2.2), where the individual results are presented. There was a significant moderate correlation between responses on the definite condition and proficiency: $r = .691$, $p = .000$ for the L2 children and $r = .672$, $p = .001$ for the L2 adults. The results for the Mann Whitney tests reported in the final column of Table 22 show that there are no significant differences in rate of scrambling between child and adult groups at any proficiency level. Comparing the three proficiency groups and adult native controls, a significant difference was observed for the L2 children (ANOVA: $df = 3$, $F = 15.513$, $p < .001$) and for the L2 adults (ANOVA: $df = 3$, $F = 17.071$, $p < .001$). Post-hoc analyses were carried out to determine within the child and adult groups, respectively, which proficiency groups differed from each other and/or from the native-speaker controls. The results of these analyses are presented in Table 23, where significant differences are highlighted.[25]

Table 23. L2 children and L2 adults (production):
Results of Games-Howell post-hoc test in definite condition

Comparison	Mean difference (%)	Significance
Child L2 Low vs. Child L2 Mid	-48.2	p = .135
Child L2 Low vs. Child L2 High	-67.3	p = .001
Child L2 Mid vs. Child L2 High	-20.0	p = .672
Child L2 Low vs. Adult natives	-70.1	p = .000
Child L2 Mid vs. Adult natives	-21.4	p = .619
Child L2 High vs. Adult natives	-1.2	p = .998
Adult L2 Low vs. Adult L2 Mid	-44.5	p = .308
Adult L2 Low vs. Adult L2 High	-71.7	p = .019
Adult L2 Mid vs. Adult L2 High	-27.6	p = .668
Adult L2 Low vs. Adult natives	-79.2	p = .005
Adult L2 Mid vs. Adult natives	-35.7	p = .372
Adult L2 High vs. Adult natives	-7.8	p = .960

The same pattern is observed for the L2 children and the L2 adults: the low proficiency groups differ significantly from the high proficiency groups and from the natives but not from the mid proficiency groups, and the mid and high proficiency groups do not significantly differ from each other or from the natives.

25 Given that equal variances could not be assumed for the child L2 groups (Levene statistic = 3.768, $p = .020$), the Games-Howell post-hoc test was employed. This is also the case for the specific indefinite condition discussed below in both the child L2 group (Levene statistic = 4.520, $p = .010$) and the adult L2 group (Levene statistic = 3.652, $p = .025$).

There is no correlation between responses in the definite condition and length of exposure for the L2 children ($r = .165$, $p = .432$), but there is a moderate correlation for the L2 adults ($r = .451$, $p = .031$). There is no correlation with age at time of testing for either group ($r = .300$, $p = .145$ for the L2 children and $r = -.002$, $p = .998$ for the L2 adults). A regression analysis shows that proficiency score is the best predictor of scrambling responses in the definite condition for both groups: $r^2 = .477$, ANOVA $F = 19.148$, $p = .000$; $t = 4.376$, $p = .000$ for the L2 children and $r^2 = .452$, ANOVA $F = 16.503$, $p = .001$; $t = 4.062$, $p = .001$ for the L2 adults.

5.4.2.1.2 Specific indefinite condition

The targetlike response in the specific indefinite condition also requires subjects to scramble the object. The results for the L2 child and adult proficiency groups are presented in Table 24. Like the native adults and the L1 children, L2 subjects sometimes produced responses which expressed a targetlike specific interpretation of the indefinite but did not actually include an indefinite NP (see §5.3.2.2 and §5.4.3.4 for examples).[26] For this reason, the number of subjects per group differs between conditions. The raw numbers are included in the table of individual results in Appendix C.

Table 24. L2 children and L2 adults (production): Average percentage targetlike (scrambled) responses in specific indefinite NP condition (organised by proficiency group)

	L2 children			L2 adults			Mann Whitney tests comparing child/adult groups
	n	%	SD	n	%	SD	
Low	10	23.3	35	7	14.3	38	Z = -.910, p = .363
Mid	6	71.7	40	6	58.3	47	Z = -.343, p = .731
High	7	91.7	14	7	85.7	38	Z = -.444, p = .657

The children and adults in the high proficiency groups and the children in the mid group all scramble relatively consistently, albeit it to a slightly lesser extent in the latter case. This again contrasts with the children and adults in the low proficiency groups, who scramble on average in approximately one fifth of items. When subjects fail to scramble; they produce one of two types of utterance, either with the object immediately preceding the verb, as in (13)-a, or with a Negation-Verb-Object order, as in (13)-b.

(13) a. Mickey gaat niet een bloem plukken
Mickey goes not a flower pick

[26] The use of such alternatives could be interpreted as an avoidance strategy or at least an indication that subjects perhaps do not know that scrambled indefinite NPs are used to express a specific meaning. Given that native speakers adult controls also produced such alternative responses, it is impossible to determine whether this is the case, however.

b. Mickey gaat niet plukken een bloem
Mickey goes not pick a flower

There is also one token with *geen*, produced by subject C25P. This contrasts with the L1 children, whose non-scrambled utterances in this condition contained the *geen*-form in just over half of the items. The indefinite determiner is sometimes realised as *één* 'one' in both scrambled and non-scrambled positions. The relationship between the form of the indefinite (*een NP* 'a NP' vs. *één NP* 'one NP') and its position will be discussed in more detail below (cf. §5.4.3.3).

There is a significant correlation between responses in this condition and proficiency for both groups (r = .721, p = .000 for the L2 children and r = .641, p = .003 for the L2 adults). As in the definite condition, the standard deviations for all groups suggest that there is some overlap between the different proficiency levels and that – with the exception of the high proficiency children – there is considerable variation within each group. With the exception of this latter group, the amount of variation is comparable for children and adults. The results for the Mann Whitney tests given in the final column of Table 24 show that within each proficiency level, children and adults do not significantly differ from each other. Comparing the three proficiency groups and the native adult controls, a significant difference was observed both for the L2 children (ANOVA: df = 3, F = 13.899, p = .000) and for the L2 adults (ANOVA: df = 2, F = 9.051, p = .000). The results of the post-hoc analyses used to determine whether the different proficiency levels differed from each other and/or from the native adults are presented in Table 25 (cf. fn. 25).

Table 25. L2 children and L2 adults (production):
Results of Games-Howell post-hoc test in specific indefinite condition

Comparison	Mean difference (%)	Significance
Child L2 Low vs. Child L2 Mid	-47.9	p = .135
Child L2 Low vs. Child L2 High	-68.8	p = .001
Child L2 Mid vs. Child L2 High	-20.2	p = .672
Child L2 Low vs. Adult natives	-70.1	p = .000
Child L2 Mid vs. Adult natives	-21.2	p = .619
Child L2 High vs. Adult natives	-1.9	p = .998
Adult L2 Low vs. Adult L2 Mid	-44.1	p = .308
Adult L2 Low vs. Adult L2 High	-71.6	p = .019
Adult L2 Mid vs. Adult L2 High	-27.4	p = .668
Adult L2 Low vs. Adult natives	-79.7	p = .005
Adult L2 Mid vs. Adult natives	-35.3	p = .372
Adult L2 High vs. Adult natives	-7.8	p = .960

The same pattern is observed as in the definite condition: for both the L2 children and the L2 adults, the low proficiency groups differ significantly from the high proficiency groups and the natives but not from the mid proficiency groups, and the mid and high proficiency groups do not significantly differ from each other or from the natives.

There is no correlation between responses in the specific indefinite condition and length of exposure for the L2 children (r = .155, p = .481), but again there is a moderate correlation for the L2 adults (r = .510, p = .022). There is no correlation between responses in this condition and age at time of testing for the L2 adults (r = .281, p = .230), but there is a very strong correlation between these two variables for the L2 children (r = .905, p = .000). This might partly be explained by the fact that age at time of testing is in turn significantly correlated with proficiency (r = .457. p = .028) (although the correlation between these two variables is almost significant for the adults, too (r = .410, p = .058)). A regression analysis shows that proficiency score is nevertheless the best predictor of scrambling responses in this condition for both groups (r^2 = .519, ANOVA F = 21.593, p = .000; t = 4.647, p = .000 for the L2 children and r^2 = .410, ANOVA F = 11.832, p = .003; t = 3.440, p = .003 for the L2 adults).

5.4.2.1.3 Non-specific indefinite condition

As was the case in the specific indefinite condition, in the non-specific condition, L2 subjects also sometimes failed to produce responses containing an indefinite NP. Rather, in a similar fashion to the native-speaker adult controls and the L1 children, they produced alternative responses which did not contain the constituents necessary to determine whether scrambling had taken place (cf. §5.3.2.2). For this reason, the number of subjects per group in this condition does not necessarily tally with those in Table 22 and Table 24. The results for the L2 subjects are presented in Table 26.

Table 26. L2 children and L2 adults (production): Average percentage targetlike (non-scrambled) responses in non-specific indefinite NP condition (organised by proficiency group)

	L2 children			L2 adults			Mann Whitney tests comparing child/adult groups
	n	%	SD	n	%	SD	
Low	8	47.9	52	7	58.3	49	Z = .293, p = .770
Mid	5	70	45	5	86.7	30	Z = -.643, p = .521
High	6	83.3	41	7	100	00	Z = -1.080, p = .280

In contrast to the other two conditions, in the non-specific indefinite condition, the targetlike response is no scrambling. There were two types of utterance which were counted as targetlike, namely those containing the suppletive form *geen* ‘no(ne)/not any’, as in (14)-a, and those without this form, but with the targetlike word order, as in (14)-b. (See §2.1.2 for why the *geen* form is counted as non-scrambled.)

(14)	a.	Ernie	gaat	geen	kikker	vangen
		Ernie	goes	no	frog	catch
	b.	Ernie	gaat	niet	een kikker	vangen
		Ernie	goes	not	a frog	catch

Although, strictly speaking, it is the response in (14)-a which should be considered targetlike – virtually all of the non-scrambled indefinites produced by native-speaker adult controls in this condition were realised as *geen* – the use of the *geen* form arguably requires the acquisition of lexical rather than syntactic-semantic knowledge. Given that it is the latter which is of interest here, utterances such as the one in (14)-b are classed as targetlike because the object is not scrambled. There were three L2 children and seven L2 adults who produced utterances with *geen*. All three L2 children (C22P, C24P and C25P) and five of the L2 adults (A3P, A8P, A9P, A13P and A18P) did so in all the items they produced and the remaining two L2 adults (A15P and A19P) did so in two out of their three responses. With two exceptions (A3P and A15P, both mid proficiency), all of these subjects belonged to the high proficiency groups.

Utterances which were classified (at least at this point in the analysis) as non-targetlike in this condition include those containing a scrambled object, as in (15)-a, and those with the Negation-Verb-Object order, as in (15)-b. The latter type of responses was found in the two low proficiency groups only, whereas the scrambled order was produced by various subjects (seven in total) in all proficiency groups except the high proficiency adults. The exact numbers are provided below in §5.4.2.2.3, where the individual results are discussed.

(15)	a.	Ernie	gaat	een kikker	niet	vangen
		Ernie	goes	a frog	not	catch
	b.	Ernie	gaat	niet	vangen	een kikker
		Ernie	goes	not	catch	a frog

The three child L2 proficiency groups do not differ significantly from each other (Kruskal-Wallis: df = 2, χ^2 = 2.588, p = .274) and neither do the adult L2 groups (Kruskal-Wallis: df = 2, χ^2 = 4.975, p = .83).[27] When the native adult controls are included in the analysis, the results are significant for both the L2 children (Kruskal-Wallis: df = 3, χ^2 = 9.440, p = .024) and the L2 adults (Kruskal-Wallis: df = 3, χ^2 = 9.819 p = .020). This means that at least one of the L2 groups differs from the natives, undoubtedly because the native-speaker adults *never* scramble in this condition. There is no correlation between

[27] The non-parametric Kruskal-Wallis test was used because the data do not conform to the assumption of equal variances which is a prerequisite for an analysis of variance (ANOVA).

proficiency scores and responses in this condition: r = .220, p = .380, for the L2 children and r = .371, p = .143 for the L2 adults. The results of the Mann Whitney test show that the child and adult groups do not significantly differ from each other within each proficiency level. There is no correlation between responses in the non-specific indefinite condition and length of exposure for either the L2 children (r = .316, p = .187) or the L2 adults (r = .292, p = .240), nor between responses in this condition and age at time of testing (r = .111, p = .652 for the L2 children and r = .376, p = .124 for the L2 adults).

5.4.2.1.4 Comparing different conditions

In order to ascertain (i) whether L2ers know the interpretive restrictions on scrambling of indefinite NPs and (ii) whether the scrambling of definite NPs and the scrambling of indefinite NPs develop in parallel, the results from the two indefinite conditions were compared with each other, as were those for the specific indefinite condition and the definite condition. Table 27 presents the first comparison and Table 28 the second. Given that the data did not meet the prerequisites for a t-test, the non-parametric Wilcoxon signed-ranks test was used to determine whether within each proficiency group, the average responses in each condition in the comparison significantly differed from each other. Statistically significant responses are highlighted.

A comparison of the two indefinite conditions involves determining whether for each proficiency group, the targetlike pattern of scrambling in the specific indefinite condition and no scrambling in the non-specific indefinite condition obtains. In order to do this, it is necessary to compare the average proportion of scrambled objects (rather than the average proportion of targetlike responses, as in Table 24 and Table 26). All utterances with the order Negation-Verb-Object are thus excluded; this is because these utterances do not involve scrambling and, consequently, including them in any calculation would mean lumping them together with non-scrambled utterances (i.e. Negation-Object-Verb, including the *geen*-forms). Whilst this is legitimate in the specific indefinite condition, because both are non-targetlike, it is not legitimate in the non-specific indefinite condition because one is targetlike (Negation-Object-Verb) and the other (Negation-Verb-Object) is not. In order to make a fair comparison between the two conditions, Negation-Verb-Object utterances were therefore excluded in the specific indefinite condition as well. This means that Table 27 and Table 28 include only those subjects who produced Negation-Object-Verb and/or Object-Negation-Verb utterances in both of the relevant conditions.

Table 27. L2 children and L2 adults (production): Comparison of average number of scrambled responses in specific indefinite condition and non-specific indefinite condition (organised by proficiency group)

	Specific indefinite condition (target: scrambled)			Non-specific indefinite condition (target: non-scrambled)			Wilcoxon signed ranks test comparing different conditions	
	n	%	SD	n	%	SD	Z	Significance
L2 children								
Low	10	23.3	35	5	23.3	43	-1.342[a]	p = .180
Mid	6	71.7	40	5	30.0	45	-1.473[a]	p = .141
High	7	91.7	14	6	16.7	41	-2.060[a]	p = .039
L2 adults								
Low	7	14.3	38	5	20.0	45	-.272[a]	p = .785
Mid	6	58.3	47	5	13.3	30	-1.841[a]	p = .066
High	7	85.7	38	7	0	00	-2.236[a]	p = .025

[a] based on positive ranks

Table 27 shows that for both the L2 children and the L2 adults, it is only in the high proficiency group that there is a statistically significant difference between responses in the two indefinite conditions. In other words, the high proficiency children and adults scramble significantly more often in the specific condition than in the non-specific condition. The difference for the mid proficiency adults is approaching significance. Subjects' responses on these two conditions did not significantly correlate with each other ($r = .405$, $p = .120$ for the L2 children and $r = .341$, $p = .233$ for the L2 adults).

The specific indefinite and definite conditions, given in Table 28, do not statistically differ from each other for any of the proficiency levels, in either the child or the adult groups. This is unsurprising given that in both groups, responses on these two conditions correlate highly ($r = .905$, $p < .001$ for the L2 children and $r = .833$, $p < .001$ for the L2 adults).

Table 28. L2 children and L2 adults (production): Comparison of average number of scrambled responses in specific indefinite condition and definite condition (organised by proficiency group)

	Specific indefinite condition (target: scrambled)			Definite condition (target: scrambled)			Wilcoxon signed ranks test comparing different conditions	
	n	%	SD	n	%	SD	Z	Significance
L2 children								
Low	10	23.3	35	12	22.1	34	-.535[b]	p = .593
Mid	6	71.7	40	6	80.0	40	-.816[b]	p = .414
High	7	91.7	14	7	88.1	31	-.447[a]	p = .655
L2 adults								
Low	7	14.3	38	9	19.3	33	-1.633[b]	p = .102
Mid	6	58.3	47	6	88.9	27	-1.604[b]	p = .109
High	7	85.7	38	8	87.5	35	.000[b]	p = 1.000

[a] based on positive ranks
[b] based on negative ranks

Before turning to the discussion of these results in §5.4.3, let us consider the results for individual subjects.

5.4.2.2 Individual results

The previous section presented the average rates of scrambling per proficiency group. There it was noted that the relatively high standard deviations suggested a certain level of variation within each proficiency level and some overlap between adjacent levels. This section considers individual subjects' response patterns in order to determine the extent of this variation and overlap. The individual data are summarised below and presented in detail in Appendix C.

5.4.2.2.1 Definite condition

Individual subjects were categorised according to whether they never scramble (indicated with "–"), sometimes scramble (" ± ") or always scramble (" + ").[28] Table 29 shows the distribution of these response patterns for each proficiency level for the child and adult groups. The first two columns provide the group

[28] The ± scrambling group includes subjects who scrambled in at least one (but not in all) items. Unfortunately, given the different number of tokens across subjects, it is not possible to employ a criterion similar to the one which will be used in the comprehension experiments in Chapter 6, where subjects are categorised according to whether they provide a targetlike response in all or all but one of the items (+ scrambling), a non-targetlike response in all or all but one of the items (- scrambling) or somewhere in-between (± scrambling). The admittedly gross division carried out here is thus intended to provide a rough indication of the subjects' responses relative to each other.

results discussed in §5.4.2.1.1 above. The cells containing the most subjects in each proficiency level is highlighted.

Table 29. L2 children and L2 adults (production):
Distribution of individual response patterns in definite condition

Group	Average rate of scrambling within group		No. of individuals with given response pattern		
	%	**SD**	–	±	+
L2 children					
Low	22.1	34	7	4	1
Mid	80.0	40	1	1	4
High	88.1	31	0	1	6
L2 adults					
Low	19.3	33	5	3	1
Mid	88.9	27	0	1	5
High	87.5	35	1	0	7

For both groups, the same relative distribution is observed: in the low proficiency groups, virtually all subjects scramble optionally or not at all, with slightly more subjects in the latter category, whereas most of the mid and high level subjects are targetlike, that is, they scramble all the time.[29]

5.4.2.2.2 Specific indefinite condition

Table 30 presents the distribution of the different response patterns for the specific indefinite condition and it repeats the group results discussed in §5.4.2.1.2. Again, the target response here is scrambling. It should be noted that these results include a number of subjects who produced just one token in this condition. These were two low proficiency children (both in the – category), one high proficiency child (+), two low proficiency adults (both –) and two high proficiency adults (both +).

---

[29] Given that the numbers in most cells are lower than 5, it is not possible to carry out, for example, a Chi-squared test to determine whether the distribution of response patterns differs per proficiency level.

Table 30. L2 children and L2 adults (production):
Distribution of individual response patterns in specific indefinite condition

Group	Average rate of scrambling within group		No. of individuals with given response pattern		
	%	SD	–	±	+
L2 children					
Low	23.3	35	6	3	1
Mid	71.7	40	1	2	3
High	91.7	14	0	2	5
L2 adults					
Low	14.3	38	6	0	1
Mid	58.3	47	1	2	3
High	85.7	38	1	0	6

Once again, the relative distribution of individuals across response patterns in each of the proficiency levels is similar across child and adult groups: in the low proficiency level, most subjects never scramble or in the child group, they do so only sometimes, whereas in the mid and high proficiency groups, most subjects scramble some or all of the time.

5.4.2.2.3 Non-specific indefinite condition

Table 31 presents the distribution of response patterns for the non-specific indefinite condition. The target in this condition is no scrambling. As in the specific indefinite condition, in the non-specific indefinite condition, too, there were a number of subjects who produced one token only. These were three children, one of each proficiency level (all in the – category) and five adults, two low, one mid and two high proficiency (all –). (Recall that for reasons given in §5.4.2.1.4 above, the individual data exclude Negation-Verb-Object sequences.)

Table 31. L2 children and L2 adults (production):
Distribution of individual response patterns in non-specific indefinite condition

Group	Average rate of scrambling within group		No. of individuals with given response pattern		
	%	SD	–	±	+
L2 children					
Low	23.3	43	3	1	1
Mid	30.0	45	3	1	1
High	16.7	41	5	0	1
L2 adults					
Low	20.0	45	4	0	1
Mid	13.3	30	4	1	0
High	0	00	7	0	0

Most subjects in each proficiency level do not scramble. Responses where the object remains in its non-scrambled position could stem from one of two sources: they may reflect low proficiency (as it does in the definite and specific indefinite conditions) or they may reflect targetlike knowledge that non-specific indefinites should not be scrambled. A comparison of individual subjects' responses on the two indefinite conditions is carried out in the next section in order to address this issue. Although the vast majority of subjects (48.8% (26/33)) do not scramble, there are some (n=7) who do produce scrambled objects in this condition some or all of the time, and they are found in all levels except the adult high proficiency group. I discuss whether these reflect non-targetlike knowledge of the interpretive constraints on scrambled objects in §5.4.3.4, which compares response patterns across conditions.

5.4.2.2.4 Subjects with other L2s

As noted in §5.4.1, some of the subjects knew other L2s besides Dutch. These were three L2 children, who were learning French at school, and 13 L2 adults, who had learned various languages prior to Dutch (mainly French and German) at school and in further education. None of the adults were exposed to these other languages before the age of 12. The details for each individual subject are given in Appendix C.

In order to determine whether knowledge of another language, and in particular knowledge of another scrambling language, such as German, and in the case of one subject, Russian, might have had an effect on subjects' responses, the individual results were examined on the basis of this variable. Two of the three (French-learning) L2 children (C20P (mid) and C25P (high)) demonstrated knowledge of scrambling and the other one (C13P (mid)) did not. Amongst the L2 adults, there were eight subjects who had knowledge of a language other than German or Russian. Four of these (A11P, A12P, A14P, A23P) were low proficiency subjects and they all failed to scramble. Of the remaining four, three (A5P (mid), A6P (high), A8P (high)) demonstrated knowledge of scrambling and the other (A7P (high)) did not. There were five L2 adults who had knowledge of German. Four (A13P (high), A15P (mid), A17P (high) and A19P (high)) demonstrated knowledge of scrambling and the other one (A4P (low)) did not. Subject A19P could also read Russian.

There appears to be no difference between those subjects who know another non-scrambling L2 and those who know another scrambling L2; in both groups, there are low, mid and high proficiency subjects and their responses on the scrambling task generally correlate with their proficiency in Dutch. Note that even if the L2ers who demonstrate knowledge of scrambling in the Dutch task were to have transferred this from their L2 German, this would not effect the claim put forward here that adult L2ers are able to acquire targetlike scrambling because as noted above, these L2ers were, on the criteria used here, all adults when they were first exposed to any of their L2s.

5.4.3 Discussion

The group and individual results presented above will now be discussed. First, in §5.4.3.1, developmental sequences are determined for the acquisition of scrambling of the three different types of direct object NP. Subsequently, in §5.4.3.2, I consider what it would mean to have an optional scrambling stage and reflect upon the prediction made for the L2 data on the basis of Schaeffer (2000). In §5.4.3.3, the indefinite data are analysed in more detail, namely in terms of whether the indefinite determiner is realised as *een* 'a' or *één* 'one'. Finally, in §5.4.3.4, the different conditions are compared with each other.

5.4.3.1 Developmental sequences

The first observation regarding the results of the production experiment is that in each condition, there are both child and adult subjects who produce targetlike responses, that is, it is possible for English-speaking children and adults to acquire direct object scrambling in Dutch. Two non-targetlike orders are produced in the two obligatory scrambling conditions: Negation-Verb-Object and Negation-Object-Verb. The Negation-Verb-Object order, which was produced by low proficiency subjects only, appears to result from the transfer of the L1 English's left-headed VP, as illustrated in (16).

(16) [$_{IP}$ Ernie [$_{I}$ will [$_{NegP}$ not [$_{VP}$ [$_{V}$ copy] the giraffe]]]]

Other studies examining the L2 acquisition of a Germanic OV language by native-speakers of a VO language have also observed the same word order pattern in the early stages of L2 development; see Clahsen (1988) on adult L1 Romance/L2 German (cf. §3.2.2.1), Felix (1977) and Molony (1977) on child L1 English/L2 Dutch, and le Roux and Wilsenach (1999) on child L1 English/L2 Afrikaans. Assuming, then, that the Negation-Verb-Object order is the result of transfer of the L1 grammar, this order must constitute the initial stage in any developmental sequence for the acquisition of scrambling by English speakers. Using the proficiency measure as an approximate guideline, it is possible to infer the developmental sequence given in Table 32 for both definite and specific indefinite NPs. The abbreviations L, M and H refer to the Low, Mid and High proficiency groups.

Table 32. Developmental sequence for L2 acquisition of scrambling of definite and specific indefinite object NPs

Stage	Description	Word order patterns produced	L2 children			L2 adults		
			L	M	H	L	M	H
I	No scrambling	Negation-Verb-Object only	2			1		
II-a	No scrambling	Negation-Object-Verb *and* Negation-Verb-Object	3			1		
II-b	No scrambling	Negation-Object-Verb only	2	1		3		1
III-a	Optional scrambling	Object-Negation-Verb *and* Negation-Object-Verb	3	3	2	2	3	
III-b	Scrambling	Object-Negation-Verb only	1	2	5	1	3	7

There appear to be three basic stages. In stage I, subjects predominantly use the Negation-Verb-Object order transferred from their L1 English, whereas in stage II, the relative order of the verb and its internal argument has switched to Negation-Object-Verb. In stage III, the targetlike scrambled order is produced. Stages I and II are split into two sub-stages, the first of which may be regarded as a period of overlap between the stage in question and the preceding stage. Stage II-a is thus characterised by the production of the non-scrambled Negation-Object-Verb order and, at the same time, the transferred Negation-Verb-Object order, and stage III-a, by the targetlike scrambled Object-Negation-Verb order and the non-scrambled Negation-Object-Verb order.

There are two low proficiency subjects whose response patterns do not fit into the sequence presented in Table 32: subject C10P produces Negation-Verb-Object and Object-Negation-Verb orders only, and subject A4P produces utterances with all three word orders. With the exception of these two subjects, all subjects who produce Negation-Verb-Object order never scramble. It seems that on their way to targetlike scrambling, L2ers pass through a stage where Negation-Object-Verb is the predominant order. This is confirmed for both L2 children and adults by two observations: first, the mid proficiency groups have more subjects who produce the Negation-Object-Verb order (some or all of the time) than the high proficiency groups, and second, the Negation-Verb-Object order is found only in the low proficiency groups.

The distribution of subjects according to developmental stage, given in Table 32, shows that the targetlike Object-Negation-Verb responses are generally – though not exclusively – produced by subjects from the highest proficiency level and the non-targetlike responses by low (and mid) proficiency subjects. There is just one high proficiency subject (A7P) who failed to scramble any definite or specific indefinite objects, but there are several low proficiency subjects who scramble. In other words, it is the case that the mid and high proficiency subjects generally scramble (some or all of the time), but it is not the case that low proficiency subjects always fail to scramble. Note, however, that it is generally the more proficient subjects within the low proficiency group who scramble; that is, within this group, we see a positive correlation between proficiency score and targetlike scrambling responses.[30] The group data reported in §5.4.2.1.1 for definite NPs and §5.4.2.1.2 for specific indefinite NPs showed no significant difference between the mid and high proficiency groups. That is, for both L2 children and adults, subjects in both proficiency levels have an average rate of scrambling comparable to native-speaker adults. This suggests that both L2 children and L2 adults reach the targetlike stage at around the same level of general proficiency.

So far the discussion has dealt with definite and specific indefinite NP objects, that is, objects which are always scrambled in the target grammar (in the contexts tested here). We now turn to objects which should not be scrambled, that is, to non-specific indefinite NPs. The developmental sequence for non-specific indefinite NPs is as given in Table 33. Recall that there are seven child L2 subjects who produced scrambled objects in the non-specific indefinite condition. These are excluded from Table 33 because, as noted above, if taken to have the targetlike specific meaning, these scrambled indefinite objects are not incompatible with the experimental scenario in this condition. Furthermore, such utterances are distributed across different proficiency levels and as such, it is not clear where they should fit into the schema below (if at all).[31] This issue is discussed further in §5.4.3.4.

---

[30] This suggests that by lowering the point at which the low and mid proficiency groups are divided (currently -0.50), the low proficiency group would mainly consist of subjects in Stage I or Stages I and II. In other words, the observation that there are low proficiency subjects in Stage III may result from the (relatively arbitrarily determined) point at which the proficiency levels are divided.

[31] It could be objected that subjects who failed to scramble in the specific indefinite condition intended the non-specific meaning. There are two reasons why this is doubtful: (i) unlike in the non-specific indefinite condition, where a specific interpretation cannot be ruled out, in the specific indefinite condition, the picture story clearly excludes a non-specific reading because there are objects which are manipulated; (ii) as noted below (§5.4.3.4), many of the subjects' responses and comments indicated that they nevertheless attributed a targetlike specific reading to the object.

Table 33. Developmental sequence for L2 acquisition of scrambling of non-specific indefinite object NPs

Stage	Description	Word order patterns produced	L2 children			L2 adults		
			L	M	H	L	M	H
I	Non-targetlike no scrambling	Negation-Verb-*een* NP only	3			2		
II	Targetlike non-scrambled order	Negation-*een* NP-Verb only	3	3	2	3	2	2
III-a	Optional use of target lexical form	*geen* NP-Verb *and* Negation-*een* NP-Verb					1	1
III-b	Targetlike non-scrambled order and lexical form	*geen* NP-Verb only			3		1	4

As in the sequence in Table 32 for definite and specific indefinite objects, Stage I for non-specific indefinites is also characterised by transfer of the L1 Negation-Verb-Object order. This is directly followed in Stage II by the targetlike Negation-Object-Verb order, but with the non-targetlike *niet een NP* form, prior to the targetlike *geen* form being acquired in Stage III. Stage III is divided into two sub-stages, the first of which may be regarded as a period of overlap between the previous and subsequent stage, that is, between Stages II and III-b. Although there are no subjects who produce both Negation-Verb-Object and Negation-Object-Verb orders in this condition, an overlap stage as in Stage II-a in Table 32 above, could in principle exist. The observation that there are subjects who produce the Negation-Verb-Object order in this condition and the Negation-Object-Verb order in another suggests that this is the case. How subjects pattern across Table 32 and Table 33 is discussed in §5.4.3.4.

5.4.3.2 On optional scrambling and Schaeffer (2000)

The developmental sequence outlined in Table 32 suggests that the acquisition of scrambling is a gradual process, where, for example, during a certain period, subjects scramble optionally.[32] This appears to be the case in L1 acquisition, too. Recall that Schaeffer observes that children pass through an optional scrambling stage; eight of the 13 L1 children tested here also scramble optionally (at rates ranging from 16.7% (1/6) to 83.3% (5/6); see §5.3.2). Optionality in the developing grammar has been linked to certain non-targetlike properties of grammar, such as the underspecification of a particular

---

[32] To speak of an optional scrambling stage only makes sense in cases where scrambling is obligatory. This holds for the instances of scrambling which are examined in this thesis, in contrast to, for example, scrambling of definite object NPs across adverbs, which is optional. It is possible that both scrambled and non-scrambled definite object NPs in the input may contribute to the existence of the optional scrambling stage observed in the data presented here.

feature (e.g. Hoekstra and Hyams 1996). On Schaeffer's approach, it is the feature [referential] which is underspecified and which leads to the optionality of scrambling. This is the consequence of L1 children's lack of a pragmatic principle which states that hearer knowledge and speaker knowledge are always separate. The existence of an L1 optional scrambling stage therefore forms a crucial part of Schaeffer's account. On the basis of this approach, it was predicted in §5.1 that given that the L2 children and adults tested here are old enough to have the pragmatic principle in question, they should, once they have knowledge of scrambling, scramble consistently, that is, an L2 optional scrambling stage should not be observed. The data presented here go against this prediction, that is, there are (both child and adult) L2ers who scramble optionally. Schaeffer (2005) counters, however, that when L2ers fail to scramble in this optional scrambling stage (and in the "no scrambling" stage which precedes it), this may be due to continuing 'interference' from their L1; in other words, the optional scrambling stage in L2 acquisition has a different source than the optional scrambling stage in L1 acquisition. As Schaeffer (2005) notes, this claim predicts that L2ers whose L1 has scrambling (e.g. German and Turkish) should not fail to scramble, that is, they should not pass through such an optional scrambling stage. This is an area for future research.

I view the 'optional scrambling' stage III-a as a transitional phase where the (L1/L2) learner is in the process of acquiring the means to generate the scrambled order.[33] Accounting for how learners progress from stage X to Y and why consecutive stages should overlap has received relatively little attention in the L2 literature (but see e.g. Gregg 1996; Schwartz and Sprouse 1994; Sorace 2000; White 2003) and I do not propose to address this issue in any great detail here. Viewed as a transitional phase, the optional production of scrambled forms could be conceptualised as the co-existence of two grammars, a scrambling grammar and a "no scrambling" grammar, which 'compete' with each other (Vainikka and Young-Scholten 1996a). When the scrambling grammar 'wins', a scrambled order is produced and when the non-scrambling grammar is the 'victor', a non-scrambled order is produced. The claim that the source of optional scrambling is the transition from one stage to another is supported by the observation that with just a few exceptions, those subjects who scramble optionally are either in the mid proficiency group or at the top end of the low proficiency group, that is, in terms of proficiency, they are somewhere between those subjects who generally fail to scramble and those who always scramble. Furthermore, optional scrambling is (generally) restricted to those types of NPs which scramble, that is, to definite and specific indefinite NPs. In other words, it is not the case that the optional scrambling exhibited

[33] Given that the focus here is on development rather than ultimate attainment, I leave open the question of whether there are L2 subjects who do not progress beyond this stage, that is, whether there are L2ers for whom scrambling remains optional.

by L2ers' is random, which might be expected if this were due to the underspecification of a [scrambling] feature (cf. §2.4), for example.

5.4.3.3 The use of *een* vs. *één*

The results presented thus far for the two indefinite conditions collapse responses where the determiner is realise as unstressed indefinite *een* 'a' and those where it is realised as cardinal/stressed indefinite *één* 'one' (cf. §2.1.2). In English, the cardinal 'one' is often used to express the specific reading of the indefinite (albeit not unambiguously specific), as exemplified in (17).

(17) Mickey didn't pick one flower

It is possible that (at least partly) as a result of transfer from their L1, L2ers use the form rather than the position of the NP to express specificity. It should be noted, however, that English-speaking L2ers of Dutch sometimes pronounce the indefinite determiner as *één* rather than *een*, and therefore, one cannot be certain that they actually distinguish these two forms. Assuming for the moment that they do – and we shall see below that there is reason to believe that on the whole, this is the case – if they were to use the form of the NP rather than its position to indicate specificity, they would thus produce utterances such as (18), where the object is realised as *één N* in non-scrambled position and the sentential stress is on either the negator *niet* 'not' or the verb.

(18) Mickey gaat niet één bloem plukken
Mickey goes not one flower pick
'Mickey isn't going to pick one flower.'

Out of all the non-scrambled utterances produced by the L2ers, 46.6% (27/58) contain an indefinite NP whose determiner is realised as *één*. (Scrambled utterances with *één* are discussed below.) Amongst the L2 adults, there are eight Negation-Verb-Object utterances with *één* and three of these are produced by one subject, and there are 20 Negation-Object-Verb utterances with *één* and three subjects produce seven of these. Amongst the L2 children, there are six Negation-Verb-Object tokens produced as *één* (out of a total of eight), produced by two subjects, and there are 11 Negation-Object-Verb utterances with *één* (out of a total of 22) and these are produced by seven subjects. Details of the distribution of *één* across individual subjects are provided in Appendix C.

All of these non-scrambled utterances are counted as non-targetlike. It might be objected, however, that this is problematic, because the non-scrambled order with *één*, exemplified in (18), is actually a true answer in the specific condition. Recall that in the story for this condition, there are six objects, three of one type and three of another. The protagonist says that for some reason, s/he is going to leave one object unmanipulated, implying that s/he will manipulate the other five and in the final picture, the subject sees the protagonist carrying out this action. The sentence in (18) is thus true because

Mickey is not going to pick one flower; rather, he is going to pick two flowers. Consequently, responses such as those in (18) could be the result of a lack of knowledge of scrambling or a(n unexpected) contrastive answer to the experimental question.

There are several reasons why the latter possibility seems unlikely. Firstly, if such a response were consistent with the experimental scenario and could in principle therefore be used as an answer, we might expect the native adults to (at least occasionally) provide such a response, but they do not. Furthermore, there is only one such response in the child L1 data. This is consistent with the idea that the L2ers' use of *één* in this condition is the result of transfer from English, rather than being intended as a response with contrastive negation. Secondly, for the utterance in (18) to be a targetlike expression of the contrastive interpretation given above (i.e. what Mickey is going to do is not pick one flower (but two)), as in the equivalent English sentence, sentential stress should be on the indefinite object and this is never the case: all L2 utterances where the non-scrambled indefinite object is realised as *één NP* rather than *een NP* have sentential stress on either the negator or the verb.[34] Furthermore, responding to the experimental question 'What is Mickey not going to do?' with a contrastive answer would be infelicitous, because such a response implicitly answers the question 'What is Mickey going to do?', that is, what Mickey is going to do is not pick one flower. Finally, the alternative responses which subjects produce (that is, the responses without the constituents necessary for scrambling) indicate that they understand the scenario and the question in the intended way and that a contrastive meaning is not construed.

Further evidence that the L2ers use the form rather than the position of the NP to indicate specificity comes from the observation that utterances where the indefinite object is realised with *één* are generally restricted to the specific condition. This furthermore suggests that the L2ers tested here make a distinction between this form and the regular, unstressed indefinite determiner *een*. There are just four instances (3.9% (4/102) across all L2 subjects) where the *één* form is used in the non-specific indefinite condition. One low proficiency child, C10P, produces one such utterance with a Negation-Verb-Object order; a low proficiency adult, A12P, produces one Negation-Object-Verb utterance with the *één* form; and the mid proficiency child C17P produces two such utterances, one with a Negation-Object-Verb order and the other with a Object-Negation-Verb order. This contrasts with the specific indefinite condition, where 53.8% (79/147) of utterances contain an NP realised with *één*.

[34] Placement of sentential stress was determined by having two raters (the present author, a fluent non-native speaker of Dutch, and a native-speaker student assistant) listen to the recordings of each of the relevant utterances. Given that there was almost complete agreement between the two raters (96.3% (26/27)), a phonetic analysis was deemed unnecessary.

Returning to the specific indefinite condition, then, let us now examine the scrambled utterances where the indefinite determiner is realised as *één*, as in (19).

(19)	Mickey	gaat	één bloem	niet	plukken
	Mickey	goes	one flower	not	pick
	'Mickey isn't going to pick one (a certain) flower.'				

It should first be noted that the native-speaker controls overwhelmingly realised the scrambled object in this condition as *één* rather than *een* (in 88.1% (37/42) of items). Such utterances were therefore classed as targetlike. Nevertheless, one might quibble that they do not necessarily provide unambiguous evidence that an L2er has knowledge of the interpretive effects of scrambling because it may be the form of the indefinite object (*één* as opposed to *een*) rather than its position which expresses the specific meaning (cf. §7.4.2 for similar discussion regarding the child L1 data). Truly unambiguous evidence of L2ers' knowledge of scrambling in the specific indefinite condition would thus be scrambled indefinite objects which are *not* realised as *één*.[35] A detailed examination of the scrambled utterances produced by L2ers reveals that such utterances exist. Amongst the L2 adults, ten out of the 32 scrambled indefinites in this condition are realised as *een* rather than *één*, and there are five subjects (out a total of 12) who realise the scrambled indefinite as *een* some or all of time. Amongst the L2 children, 27 of the 57 scrambled indefinites in this condition are realised as *een* rather than *één* and there are ten subjects (out of a total of 16) who realise the scrambled indefinite as *een* some or all of the time. There is thus evidence that (at least some) L2ers use the scrambled order without the *één* form to express the specific reading of the indefinite object NP. The fact that many of these objects are realised as *één* could just as well be an indication of targetlike knowledge rather than L1 transfer, because as noted above, this is the preferred realisation of the scrambled object in this condition by the native-speaker controls.

---

[35] Note, however, that as is the case for English 'one', it is not strictly true that Dutch *één* is unambiguously specific: the non-scrambled indefinite object in (i) is likely to be interpreted specifically, but not necessarily. In other words, the sentence in (i) could mean that the boy threw two different balls (in Dutch and in English).

(i)	De jongen	heeft	twee keer	één bal	gegooid
	the boy	has	two times	one ball	thrown
	'The boy threw one ball twice.'				

Nevertheless, given the tendency to interpret 'one' and *één* as referring to a specific entity, examining the data both with and without *één* allows us to tease apart L1 transfer and targetlike L2 knowledge. It also means that we can determine whether L2ers exhibit targetlike knowledge under the strictest criteria.

In this section, it has been established that the intended meaning of non-scrambled utterances with *één* is unlikely to be one of contrastive negation, and that rather, such utterances are due to a lack of knowledge of scrambling combined with L1 transfer. Furthermore, although realising the indefinite determiner as *één* is probably the most targetlike way of realising the scrambled object, there are several L2 subjects who produce scrambled objects with *een* rather than *één*, providing clear indication that it is the position of the indefinite NP rather than its form which is used to express the object's specific interpretation.

5.4.3.4 Comparing conditions

As noted above in §5.4.2.1.4, by comparing the subjects' response patterns across conditions, it will be possible to determine whether (i) L2ers know the interpretive restrictions on scrambling of indefinite NPs and (ii) the scrambling of definite NPs and the scrambling of indefinite NPs develop in parallel. The results presented there indicate that at a group level, a statistically significant difference between the specific and non-specific indefinite conditions was observed for the two high proficiency groups only; this difference approached significance for the mid-level adults. There was no significant difference between the specific indefinite and definite conditions for any group. A closer examination of the individual results (cf. §5.4.2.2 and Appendix C) is necessary in order to determine to what extent these (lack of) differences hold within and between subjects, as opposed to within and between proficiency levels.

Let us first consider whether the L2ers know the interpretive constraints on scrambling of indefinite NPs by comparing individual response patterns across the non-specific and specific indefinite conditions. There are six L2 children and eight L2 adults who fail to produce indefinite objects in one or both conditions; these are excluded from the following discussion. Amongst the remaining 19 child L2 and 15 adult L2 subjects, the patterns given in Table 34 are observed. The subjects who produced one item only in one of the two conditions are indicated with an asterisk. As in previous tables, the symbol "–" is used to indicate no scrambling (Negation-Object-Verb and/or Negation-Verb-Object order), " ± " gives subjects who sometimes scramble, and " + " subjects who scramble all the time. Given that optional and obligatory scrambling are analysed as being part of the same developmental stage (cf. §5.4.3.1), the latter two categories are combined for the purpose of this analysis.

Table 34. L2 children and L2 adults (production): Individual response patterns across specific indefinite and non-specific indefinite conditions

Response pattern			No. of subjects with each pattern			
			L2 children		L2 adults	
Pattern	Spec. indef.	Non-spec. indef.	n	Which subjects?	n	Which subjects?
a	–	–	6	Low : *C3P, *C4P, C5P, *C7P, C10P Mid : *C13P	5	Low : *A1P, *A20P, A23P Mid : A3P High : A7P
b	± / +	–	8	Low : C1P Mid : C14P C20P High : C21P, C22P, *C23P, C25P, C25P	8	Low : *A2P Mid : A10P, A15P High : *A9P, A13P, *A17P, *A18P, A19P
c	–	± / +	0	---	0	---
d	± / +	± / +	5	Low : C9P, C18P Mid : C17P High : C16P, C19P	2	Low : A21P Mid : A22P

The 11 subjects exhibiting pattern (a) in Table 34 do not produce any scrambled indefinite objects, which suggests that they do not know that the targetlike means of expressing the specific meaning of the indefinite object NP is to scramble the object. Note that the responses provided by these subjects, who are mostly at the low proficiency level, indicate that their non-scrambled responses cannot be attributed to their having misinterpreted the experimental scenario. These are exemplified in (20).

(20) a. Hij gaat één rode bloem achterlaten (A3P)
he goes one red flower behind-leave
'He's going to leave one red flower behind.'

b. Hij zal niet die laatste rode bloem pakken (A7P)
he shall not that last red flower take
'He won't take that last red flower.'

c. Hij gaat niet alle bloemen plukken (C16P)
he goes not all flowers pick
'He's not going to pick all the flowers.'

The 16 subjects under pattern (b) produce scrambled objects in the specific indefinite condition and non-scrambled objects in the non-specific indefinite condition, that is, they exhibit knowledge of the interpretive constraints on the scrambling of indefinite object NPs. With two exceptions, these subjects are all at the mid or high proficiency level.

Of the remaining seven subjects who scramble in both conditions (pattern (d)), four do so consistently (i.e. 100% in both conditions). These four

(C16P, C18P, C19P and A21P) always scramble in the definite condition, too (see below for further discussion). Is it possible that these subjects (and possibly the other three) are in a stage in the development of scrambling where objects are scrambled across the board? I consider this unlikely for two reasons. First, these subjects belong to different proficiency groups, that is, the scrambling behaviour does not appear to characterise a particular level of proficiency, which one might expect if there were such a 'scrambling-across-the-board' stage. Second, although the production of a scrambled object is unexpected in the non-specific condition, it is not, as noted in the discussion of the child L1 results, strictly speaking ungrammatical (cf. §5.3.3). Recall that this scenario was designed such that there were numerous objects which were difficult to distinguish from each other in order to discourage subjects singling out any particular object. This does not however preclude subjects from doing so. Consequently, it is possible that the specific reading of the indefinite was actually the one intended by these subjects. Note furthermore that there is one response pattern across the two indefinite conditions which did *not* occur, namely scrambling in the non-specific indefinite condition and no scrambling in the specific indefinite condition (pattern (c)). In other words, it is only those subjects who scramble where they are supposed to (i.e. in the specific indefinite condition) who scramble in the non-specific condition; subjects do not scramble in the non-specific condition and no other condition. On the basis of these production data alone, it is unfortunately not possible to determine whether, when the subjects under pattern (d) scramble in the specific indefinite condition, they do so for the right reasons (given that they make no contrast with the non-specific indefinite condition, which is what is used to indicate targetlike knowledge for the other subjects). Several of these subjects did, however, take part in two of the comprehension experiments reported on in Chapter 6, and so in Chapter 7, a comparison of the production and comprehension data will shed some light on this issue.

We now turn to the comparison of the definite and specific indefinite conditions and the question of whether the scrambling of these two types of object NPs develops in parallel. There were five subjects (C2P, C8P, A8P, A11P and A12P) for whom there were no data in the specific indefinite condition; these are therefore excluded from discussion here. The patterns observed in the data from the remaining subjects are given in Table 35.

Table 35. L2 children and L2 adults (production): Individual response patterns across definite and specific indefinite conditions

Response pattern			No. of subjects with each pattern			
			L2 children		L2 adults	
Pattern	**Def.**	**Spec. indef.**	**n**	**Which subjects?**	**n**	**Which subjects?**
a	–	–	6	Low : *C3P, C4P, C5P, C6P, *C7P Mid : C13P	4	Low : A1P, *A16P, A23P High : A7P
b	± / +	–	1	Low : C10P	4	Low : A4P, *A14P, A20P Mid : A3P
c	–	± / +	0	---	0	---
d	± / +	± / +	16	Low : C1P, C9P, C12P, C18P Mid : C11P, C14P, C17P, C19P, C20P High: C15P, C16P, C21P, C22P, C23P, C24P, C25P	12	Low : A21P Mid : A2P, A5P, A10P, A15P, A22P High : A6P, *A9P, A13P, A17P, *A18P, A19P,

The majority of child L2 and adult L2 subjects scramble in both conditions some or all of the time (pattern (d)). Most of the subjects with this pattern are at the mid and high proficiency level. There are also a handful of mainly low proficiency subjects who fail to scramble in both conditions (pattern (a)). The pattern observed in (b) and the absence of any subjects showing pattern (c) suggest that definite object NPs are scrambled before specific indefinite NPs, that is, while there are subjects who scramble definites and fail to scramble indefinites, there are no subjects who scramble indefinites and fail to scramble definites.[36] This claim is supported by the observation that the subjects with pattern (b) are all at the low or mid proficiency level. There are two factors which might be at play here. First, it is possible that the scrambling of definites before indefinites may reflect the relative frequency of these objects in scrambled position in the input.[37] Second, that a distinction is made between scrambling of these two types of objects might be considered unsurprising,

---

[36] Comparing the rate of scrambling in both conditions for those subjects who sometimes scramble in both conditions (grouped under (d)) reveals a roughly even split between subjects who scramble more in the definite condition than in the specific indefinite condition and vice versa. Such a comparison only tells us which type of objects are scrambled more frequently *once* they are scrambled at all, however. The crucial observation here is the absence of pattern (c).

[37] As noted in Chapter 3, there are unfortunately no data available on the frequency of scrambled objects in Dutch. Nevertheless, most native-speaker linguists agree that indefinites generally appear in non-scrambled rather than scrambled position.

given that in certain respects, they are different. More specifically, scrambling of (anaphoric) definite NPs across negation is the result of the interaction between negation and focus, whereas the scrambling of indefinite NPs induces a clear difference in meaning, that is, scrambled indefinites are of a different type from non-scrambled indefinites.[38] This observation alone cannot, however, explain why scrambling of definites might precede scrambling of indefinites. What is needed is an independent reason for why interpretive differences relating to focus should be acquired before interpretive differences relating to semantic types and accommodation/discourse integration. Teasing apart the various parts of discourse/pragmatic knowledge is an area for future research.

5.5 Cross-group comparisons

The focus of this thesis, as outlined in Chapter 1, is a comparison of child L2 acquisition with adult L2 acquisition, on the one hand, and with child L1 acquisition, on the other. In Chapter 3, two predictions were formulated for each of these comparisons. This section evaluates whether these predictions, repeated at the beginning of this chapter, are borne out in the production data which were presented above.

5.5.1 Child L2 ~ Adult L2

For the comparison between child L2 and adult L2 acquisition, the following prediction was made:

PREDICTION ➊. CHILD L2 ~ ADULT L2: DEVELOPMENTAL SEQUENCES
L2 children and L2 adults will pass through the same developmental stages.

An independent proficiency measure was used as a basis for the child L2 ~ adult L2 comparison; this ensured that L2ers compared with each other were at approximately the same proficiency level in Dutch. The proficiency data were also used to derive a developmental sequence from the available cross-sectional data. It was observed that L2 children and L2 adults pass through the same developmental stages in their acquisition of scrambling. Low proficiency L2ers in both groups, including those whose level in Dutch was so low they were unable to complete the proficiency task, generally either failed to scramble definite or specific indefinite object NPs or only scrambled them sometimes. They also produced the Negation-Verb-Object order transferred from their L1 English.

[38] Scrambling indefinite NP objects serves to indicate that the indefinite is either a free variable (Van Geenhoven 1998) or a quantifier (de Hoop 1992), rather than being incorporated into the verb. See §2.4 and §2.5 for more details.

(21)

Stage	Description	Word order patterns produced
I	No scrambling	Negation-Verb-Object only
II-a	No scrambling	Negation-Object-Verb *and* Negation-Verb-Object
II-b	No scrambling	Negation-Object-Verb only
III-a	Optional scrambling	Object-Negation-Verb *and* Negation-Object-Verb
III-b	Scrambling	Object-Negation-Verb only

In these two conditions, the mid and high proficiency levels in both groups scrambled sometimes or always and the high proficiency subjects generally always scrambled. There were no significant differences between the L2 child and L2 adult groups in any of the proficiency levels in any of the conditions.

L2 children and L2 adults were also observed to pattern alike in the acquisition of the non-scrambling of non-specific indefinite objects, as given in (22). There was, however, one minor difference between the two groups here, namely that the only subjects in stage III-a were two L2 adults. This could be because L2 children do not pass through such a stage, or it could be because there are relatively few child L2 subjects (n=3) who produce the *geen* form at all (cf. seven L2 adults who produce this form some or all of the time). More data are needed to address this issue.

(22)

Stage	Description	Word order patterns produced
I	Non-targetlike no scrambling	Negation-Verb-*een* NP only
II	Targetlike non-scrambled order	Negation-*een* NP-Verb only
III-a	Optional use of target lexical form	*geen* NP-Verb *and* Negation-*een* NP-Verb
III-b	Targetlike non-scrambled order and lexical form	*geen* NP-Verb

With the exception of this one minor difference, then, prediction ❶ is thus confirmed: L2 children and L2 adults pass through the same developmental stages in the production of scrambled objects.

5.5.2 Child L2 ~ Child L1

For the child L2 ~ child L1 comparison, the following prediction was made:

PREDICTION ❸. CHILD L2 ~ CHILD L1: DEVELOPMENTAL SEQUENCES

As a result of L1 transfer, L2 children will pass through a different developmental sequence from L1 children.

On the basis of the previous literature and the data collected from L1 children reported on here, the following developmental sequence was hypothesised for the L1 acquisition of scrambling (of definite and specific indefinite object NPs):

(23)	**Stage**	**Description**	**Word order patterns produced**
	I	No scrambling	Negation-Object-Verb only
	II-a	Optional scrambling	Object-Negation-Verb *and* Negation-Object-Verb
	II-b	Scrambling	Object-Negation-Verb only

This sequence differs from the one given above in 0 for L2 children in that in (23) there is no initial stage where the order Negation-Verb-Object is produced. This is expected if the latter order is the result of L1 transfer. The presence of this stage in the L2 developmental sequence for non-specific indefinite NP objects in (22) is also different from the sequence proposed for L1 acquisition, given in (24). Recall that there are no L2 children in stage III-a in (22), that is, there are no children who alternately produce the *geen* form *and* the *niet* +*een* form. As noted above, this could be due to the low number of children who produce the *geen* form at all. There is one L1 child in this stage, stage II-a in (24). The observation that there is just one child in this stage could be because like the L2 children, who generally fail to produce the *geen* form, the L1 children generally fail to produce the *niet* +*een* form. More data are therefore needed to establish whether this is indeed a stage in the L1 acquisition of (the non-scrambling of) non-specific indefinite NP objects.

(24)	**Stage**	**Description**	**Word order patterns produced**
	I	Targetlike non-scrambled order	Negation-*een N*-Verb only
	II-a	Optional use of target lexical form	*geen N*-Verb *and* Negation-*een N*-Verb
	II-b	Targetlike non-scrambled order and form	*geen N*-Verb only

Prediction ❸ is thus confirmed: L2 children pass through a different developmental sequence from L1 children in the production of scrambled definite and specific indefinite objects. Note, however, that aside from the first L1 transfer stage, the remaining stages are generally the same for both L2 and L1 children, and L2 adults.

5.6 Summary

This chapter reported an elicited production experiment on scrambling of three different types of objects – definite NPs and specific and non-specific indefinite NPs – in Dutch. Three different learner groups were tested: L1 children, L2 children and L2 adults. The child L1 data indicated that most 5-year-olds consistently scramble definite NP objects and they also make a targetlike distinction between specific and non-specific indefinite NP objects, scrambling the former and leaving the latter non-scrambled. For those subjects for whom the relevant data were available, it was observed that English-speaking L2 children and adults also made the relevant distinction between the two types of indefinite NP objects. The data suggest that definite NP objects are scrambled earlier than indefinite NP objects. Developmental sequences were established for all three groups and it was found that L2 children and L2 adults pass through the same stages, with an initial L1 transfer stage which differs from L1 children, thereby confirming the predictions which were tested.

In Chapter 6, we turn to the comprehension of scrambled (indefinite) objects.

Chapter 6

The comprehension of scrambled objects

Introduction

The previous chapter investigated the development of scrambling in L1 children, L2 children and L2 adults on the basis of production data. The 5-year-old L1 children tested there were found to scramble both definite and specific indefinite objects relatively consistently, and in their scrambling behaviour, they made a clear distinction between specific indefinite objects, on the one hand, and non-specific indefinite objects, on the other. The high (and some mid) proficiency L2 children and L2 adults were found to pattern similarly, that is, in their production, they demonstrated knowledge of the interpretive constraints on scrambling. In this chapter, a series of comprehension tasks are employed to examine how these three groups of learners *interpret* scrambled objects. Previous research suggests that L1 development in this domain is gradual and that adultlike levels of comprehension are attained only very late, much later than when targetlike production is in place. New L1 data are collected to further investigate this observation. As in the preceding chapter, extending the domain of investigation to L2 acquisition, and in particular to L2 children, will demonstrate how L2 data can be used to evaluate the validity of proposals put forward to account for L1 development.

Two truth value judgement tasks were carried out with all three learner groups, one on scrambling across negation and one on scrambling across *twee keer* 'twice'. For the L1 children, the results confirm previous research: until age 9, L1 children's responses significantly differ from those of native adults. On the *twee keer* task, the L2 children's data broadly follow the same gradual, age-related pattern of development and although there is a positive correlation between targetlike responses and proficiency, the high proficiency children remain significantly different from the native adults. The L2 adult data show a similar positive correlation between proficiency and targetlike responses on this task, but for this group, the high proficiency subjects do not significantly differ from the native adults. These results suggest that the same age-related effect in L1 acquisition is also at play in child L2 acquisition. The L2 results on the negation task contrast sharply with those on the *twee keer* task: there is no correlation between targetlike responses and age for the L2 children, or between targetlike responses and proficiency for either of the two groups. It is argued that the difference in these two sets of results is due to the effect of a potential methodological problem interacting with L1 transfer in the negation task.

A third task, an act-out task on scrambling across *twee keer*, was also carried out with a subset of the L2 adults. The results support the findings of the truth value judgement task in that the high proficiency adults make the targetlike distinction between scrambled and non-scrambled indefinite objects, interpreting the former specifically and the latter non-specifically. Taken together, these results indicate that both English-speaking adult and child L2ers are able to overcome the poverty-of-the-stimulus problem which the acquisition of the interpretive constraints on scrambled indefinites in Dutch represents.

The chapter is organised as follows. Section 6.1 briefly recapitulates the previous findings on the comprehension of scrambled objects and spells out the predictions which will be tested here in more detail. Section 6.2 details the test subjects and sections 6.3 through 6.5 present the methodology and results of the three tasks employed here, namely the truth value judgement task on scrambling across negation in §6.3, the truth value judgement task on scrambling across the frequency adverb *twee keer* 'twice' in §6.4, and the act-out task, also on scrambling across *twee keer*, in §6.5. The chapter is completed with a comparison of the various tasks and learner groups in §6.6, and a summary and conclusion follow in §6.7.

6.1 Recap: Previous findings and predictions

In Chapter 3, it was argued that for the English-speaking L2er of Dutch, the acquisition of the specific interpretation of scrambled indefinite objects constitutes a poverty-of-the-stimulus problem. English-speaking L2ers can neither induce the interpretive restrictions on scrambling from the input nor can they transfer them from their L1, nor are they taught these restrictions in the foreign language classroom. Consequently, if they come to know these restrictions, their interlanguage grammars must be constrained in the same way as native grammars.

There have been very few studies on the L2 acquisition of scrambling and those which exist have concentrated on the acceptability of scrambled orders rather than the interpretive constraints on scrambling (Neeleman and Weerman 1997). Recent research on other properties of language at the syntax-semantics interface (e.g. Dekydtspotter *et al.* 1999; Marsden 2004) has established that in the L2 acquisition of a restricted scope interpretation associated with a particular form, (adult) L2ers whose L1 has only one, scopally ambiguous form, initially transfer this ambiguity to their interlanguage grammar. Eventually, however, they are able to overcome the poverty-of-the-stimulus problem to acquire the scopal restrictions present in the target language (TL). The present study seeks to build on and contribute to this body of research in two ways: first, by examining a comparable acquisition scenario in a different L1/TL pair, and second, by investigating the syntax-semantics interface in child L2 acquisition as well as in adult L2 acquisition. Furthermore, unlike the previous studies on the L2 acquisition of scrambling, the present study deals with the interpretive constraints on scrambling rather than the acceptability of the scrambled form.

This thesis compares child L2 acquisition with adult L2 acquisition, on the one hand, and with child L1 acquisition, on the other. In the case of the production of scrambled objects, reported in Chapter 5, these comparisons proved to be rather straightforward. With respect to the comprehension of scrambled objects, however, previous research in L1 acquisition shows that L1 children as old as 12 sometimes fail to assign the targetlike, specific interpretation to a scrambled indefinite object NP; the three-way comparison in this situation may therefore be rather more complicated. Krämer (2000) found that 4- to 7-year-old L1 Dutch children generally rejected sentences such as the one in (1)-a in a context where they were true, and when, in an act-out task, they were asked to carry out the action in (2)-a, they manipulated two objects instead of one. In other words, L1 children pass through a developmental stage where they assign a non-specific interpretation to scrambled indefinite objects, interpreting them in the same way as they would the non-specific, non-scrambled indefinite objects in (1)-b and (2)-b.

(1) a. De jongen heeft een vis niet gevangen
the boy has a fish not caught
'The boy didn't catch a (certain) fish.'

b. De jongen heeft geen vis gevangen
the boy has no fish caught
'The boy didn't catch a(ny) fish.'

(2) a. Je mag een potje twee keer omdraaien
you may a jar two times over-turn
'You may turn over a (certain) jar twice.'

b. Je mag twee keer een potje omdraaien
you may two times a jar over-turn
'You may turn over a(ny) jar twice.'

Philip (2003) corroborated this pattern of late acquisition. Using sentences of the (2)-a type in a truth value judgement task, he observed that 44.2% (19/43) of the 12-year-olds in his study failed to reject the non-specific reading for the scrambled indefinite object. Philip claims that this gradual and slow acquisition follows from the (optional) application of a domain-general cognitive rule, which results in acquisition proceeding on the basis of indirect negative evidence only. In Chapter 3, Philip's approach was found to be problematic, however, partly because it is based upon a rather implausible acquisition scenario (see §3.3.1.2 for details). Philip nevertheless makes an interesting claim regarding L2 acquisition, namely that the slow and gradual developmental sequence observed in L1 acquisition is also expected in (adult) L2 acquisition, because, on his account, both involve the application of domain-general knowledge rather than domain-specific (i.e. linguistic) knowledge. The task facing the English-speaking L2er is complicated by the fact that English does not encode the relevant (specificity) distinction, crucial to Philip's approach, and this, too, must be acquired. This means that, on Philip's account, in their acquisition of the interpretive constraints on scrambling, English-speaking adult and child L2ers are expected to take at least as long as L1 Dutch children. In other words, we should not find any English-speaking L2er who acquires the specific meaning of scrambled indefinites more quickly than an L1 Dutch child. This prediction will be spelled out in more detail in §6.4.3.3. There, 'speed of acquisition' is understood as years of exposure.

Krämer (2000) tested L1 Dutch children's comprehension of scrambled indefinites, for which she offers a semantic-pragmatic account. In what she dubs the 'Non-Integration Hypothesis', she claims that the semantic properties of scrambled indefinites ('free variables' on her approach) require discourse integration, and hence those L1 children with limited discourse integration will interpret scrambled indefinites as though they were not scrambled, that is, as predicative indefinites.[1] As noted in Chapter 3 (§3.3.1.1), Krämer ultimately remains agnostic with respect to the underlying cause of children's limited discourse integration. The suggestions she makes, however, are largely based on children's limited processing abilities, that is, processing limitations might prevent children from combining information from different sources (semantic, pragmatic, syntactic) and this, in turn, would lead to a failure to interpret the scrambled indefinite in the targetlike fashion. Processing limitations may also play a role in (child and adult) L2 acquisition. These might be of (at least) two types, namely those which stem from cognitive immaturity, which are only valid for L2 children, and those caused by a low L2 proficiency level, which are relevant to both groups. Child and adult L2ers would therefore be expected to differ in these terms. Given that for adult L2ers, any processing limitations can only be of the second type, it is only low-level L2 adults who could plausibly be expected to have the processing limitations which may underlie problems with discourse integration, whereas such limitations would be more pervasive amongst the L2 children. We may also expect these limitations to be related to chronological age, that is, older children will have better discourse integration skills than younger children. The research on discourse integration reviewed by Krämer (cf. §3.3.1.1) suggests that children as old as age eight and beyond fail to integrate discourse in an adultlike fashion. This is in line with both Krämer's and Philip's empirical results.

[1] Whether Krämer's account can be reconciled with the production data presented in Chapter 5 is addressed in Chapter 7.

It was noted in Chapter 3 that L1 children acquiring any language (not just Dutch) are expected to experience problems with discourse integration, that is, such problems are not specific to a given language and thus we may expect them to persist until (at least) around age eight for all children. This somewhat complicated the child L2 ~ adult L2 comparison, such that prediction ❶ was supplemented with prediction ❷.

PREDICTION ❶. CHILD L2 ~ ADULT L2: DEVELOPMENTAL SEQUENCES
L2 children and L2 adults will pass through the same developmental sequence.

PREDICTION ❷. CHILD L2 ~ ADULT L2: DISCOURSE/PRAGMATIC FACTORS
a. L2 children within the same age range as L1 children with limited discourse integration will differ from L2 adults in their interpretation of scrambled indefinite objects.
b. L2 children outside of this age range should pattern similarly to L2 adults.

And for the child L2 ~ child L1 comparison, prediction ❸ was supplemented with prediction ❹.

PREDICTION ❸. CHILD L2 ~ CHILD L1: DEVELOPMENTAL SEQUENCES
As a result of L1 transfer, L2 children will pass through a different developmental sequence from L1 children.

PREDICTION ❹. CHILD L2 ~ CHILD L1: DISCOURSE/PRAGMATIC FACTORS
L2 children within the same age range as L1 children with limited discourse integration will behave similarly to L1 children in their interpretation of scrambled indefinite objects.

I will now spell out these predictions in more detail. Prediction ❷ concerns the comparison between L2 children and L2 adults. It states (in ❷-a) that differences between these two groups are expected where age at time of testing is a relevant factor for the L2 children. When this is not the case (in ❷-b), similarities are expected between these two groups, as per prediction ❶. Implicit in this prediction is the assumption that child L2 acquisition is constrained by UG (see §1.2.1.2 for relevant discussion) and thus we expect to find English-speaking L2 children who have targetlike knowledge of the interpretive constraints on scrambled indefinite objects in Dutch. If adult L2 acquisition is also constrained by UG, targetlike L2 adults will also be found. The production data presented in Chapter 5 demonstrated that this was the case. To say that adult L2 acquisition is UG-constrained does not of course mean that all the adult L2ers tested here should be targetlike, given that the subjects come from a range of proficiency levels. In the production experiment, knowledge of scrambling was observed to develop with proficiency. It is thus expected that the high proficiency subjects will be more likely to be targetlike. On prediction ❷-b, then, it is expected that for both the older L2 children and the L2 adults, the proportion of targetlike responses will increase as a function of proficiency, that is, as per prediction ❶, these two groups will pass through the same developmental sequence.

Prediction ❹ states that those L2 children who are of the same age as those L1 children who (are claimed to) have limited discourse integration should behave similarly to the L1 children, that is, they should fail to assign a specific interpretation to a scrambled indefinite NP. Note, however, that transfer from L1 English could also lead to a failure to assign a specific interpretation to scrambled indefinites, that is, on prediction ❸, L2 children are also expected to provide non-targetlike responses as a result of transfer.[2] Despite the overlap between the predictions made by L1 transfer and limited discourse integration, it should, nevertheless, be possible to tease these two factors apart, namely in the following way. Consider limited discourse integration as the crucial factor first: on the assumption that discourse integration develops with age, then on this scenario, a correlation between age at time of testing and targetlike scrambling responses would be expected, irrespective of proficiency level. Crucially, if the age-effect in child L1 acquisition is also at play in child L2 acquisition, younger L2 children of even the highest proficiency level will fail to assign the targetlike specific interpretation to scrambled indefinite objects. Next, consider L1 transfer as the crucial factor: in this case, a correlation is expected between L2 proficiency and targetlike scrambling responses, that is, as L2ers become increasingly proficient in the L2, they should become more targetlike. Importantly, this should be possible at any age. If both of these factors, that is, if (age-dependent) limited discourse integration *and* L1 transfer are at play, it would be expected that only older, high proficiency child L2ers would be more likely to have targetlike responses.

To summarise: Taking these predictions together, the following situation is expected. As they get older, both L1 children and L2 children will have more targetlike scrambling responses, but as a result of limited discourse integration, the high proficiency L2 children in the relevant age range will be significantly less targetlike than the high proficiency L2 adults. High proficiency L2 children beyond this age range should pattern like high proficiency L2 adults, that is, they are expected to be more targetlike.

Before presenting the data addressing these predictions, details of the test subjects in each of the three learner groups are provided.

[2] The situation with scrambling across negation is slightly more complex, however; this is explored in more detail in §6.3.3.

6.2 Subjects

Three different sets of learners – L1 children, L2 children and L2 adults – plus native Dutch adults were tested using the same two truth value judgement tasks. A subset of the L2 adults were also tested using a third (act-out) task. Twelve L2 children and eight L2 adults had already taken part in the production experiment. (The results from both tasks are compared in Chapter 7.)

6.2.1 L1 children and adults

There were six groups of (monolingual) L1 Dutch children, ranging from seven to 13 years old. Data were also collected from 25 native-speaker adult controls, most of whom were students (but not in linguistics) at Utrecht University. A subset of this total number of native adults participated in each of the experiments, with five taking part in all three. Details of all L1 groups are provided in Table 36.

Table 36. L1 children and adults (comprehension): Biodata

Group	n	Age (range)
7-year-olds	14	7;0-7;10
8-year-olds	15	8;0-8;11
9-year-olds	15	9;0-9;8
11-year-olds	15	10;10-11;11
12-year-olds	14	12;0-12;10
13-year-olds	25	13;1-13;11
Adults	25	18-43

6.2.2 L2 children and adults

The L2 subjects were recruited from six different international schools in The Netherlands where English is the language of instruction, and via personal contacts as well as through advertising on a mailing list for English speakers in The Netherlands. All child and all but two adult (A26C and A20C) subjects had received some Dutch language instruction, either at the international schools they attended or in foreign language courses, and they had had varying amounts of contact with Dutch. The children received between 40 and 120 minutes of Dutch language instruction per week. All of the subjects had been raised in monolingual English-speaking families. Some of the adults had learned one or more foreign languages at school or university, but none had started before the age of 12; for those subjects who were (relatively) fluent in another language, Dutch was still the L2 in which they were the most proficient (cf. fn. 17 in Chapter 5). All (child and adult) subjects regularly used their L1. Individual biodata, including knowledge of other languages and amount of contact with Dutch, are provided in Appendix D.

There were 36 L2 children. As per the definition of child L2 acquisition given in Chapter 1, their age at first exposure was between 4;0 and 7;3 (mean = 5;7; SD = 1;0). Their age at time of testing ranged from 7;3 to 14;11 (mean = 10;4; SD = 2;3) and their length of exposure from 0;7 to 10;5 (mean = 4;9; SD = 2;11). Most of the children came from middle class families who had moved to The Netherlands for professional reasons. There were 37 L2 adults (as defined in Chapter 1). Their age of first exposure was between 8;0 and 43;0 (mean = 19;7; SD = 10;1) and their age at time of testing between 9;8 and 50 (mean = 24;0; SD = 13;4). They had been resident in The Netherlands for between two months and 25 years (mean = 4;4; SD = 5;3). The vast majority had attended higher education.

Several other children were tested but they had to be excluded for the following reasons. One was raised speaking another language (Afrikaans) in addition to English, and there were six whose Dutch proficiency level was too low to complete the task. There were also 12 other children who were exposed to Dutch before the age of four. Finally, two children were excluded during the experiment because their answers indicated that they had a response bias.[3]

6.3 Task I: Truth value judgement task with negation

The purpose of this experiment is to determine whether language learners assign a targetlike interpretation (specific vs. non-specific) to an indefinite object NP scrambled across negation, as in (1)-a above.

6.3.1 Method

This experiment is a replication of Krämer's (2000) truth value judgement task (cf. §3.3.1.1).

6.3.1.1 Materials

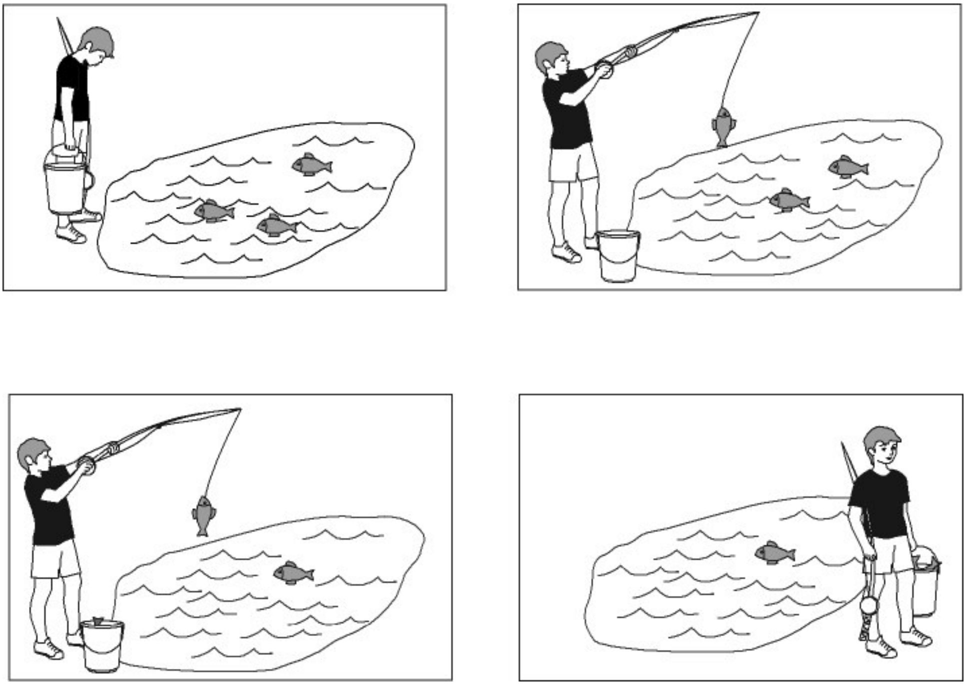
Subjects were presented with a series of four pictures which told a story about a child and three objects. In the course of the story, two of these three objects are manipulated, leaving one remaining unmanipulated object in the final picture. This is illustrated in (3). The story should be read from top left to top right, then from bottom left to bottom right.

In this example, the boy catches two out of three fish, leaving the third fish behind. As a description of this context, the scrambled sentence in (1)-a, repeated here in (3)-a, is correct, because there is a fish which the boy did not catch, namely the one left in the pond. This contrasts with the non-scrambled sentence in (1)-b, repeated here in (3)-b, which is incorrect, because it is not true that the boy did not catch any fish. The non-scrambled order was not used in the present study. There were two reasons for this. First, given that the youngest of Krämer's L1 subjects (4-year-olds) were already 100% targetlike on

---

[3] One subject consistently answered 'yes' and the other 'no'.

this order, it seemed unnecessary to test the older L1 subjects on this order here. Second, L2ers usually receive instruction on *geen* 'no(ne)/not any'. The non-scrambled form is included in the task on scrambling across *twee keer* (cf. §6.4), however, as such considerations do not play a role there.



(3) Example scenario for experimental item in truth value judgement task on scrambling across negation

Picture 1 Dit is een jongen. En dit zijn vissen. Hij wil ze denk ik gaan vangen.
'This is a boy. And these are fish. I think he wants to catch them.'

Picture 2 Hier vangt ie een vis.
'Here he's catching a fish.'

Picture 3 En hier vangt ie een vis.
'And here he's catching a fish.'

Picture 4 En nu gaat ie weer weg.
'And now he's leaving.'

Puppet:	a.	De jongen	heeft	een vis	niet	gevangen	[target: acceptance]
		the boy	has	a fish	not	caught	
		'The boy didn't catch a (certain) fish.'					
	b.	De jongen	heeft	geen vis	gevangen		[target: rejection]
		the boy	has	not-a fish	caught		
		The boy didn't catch a(ny) fish.'					

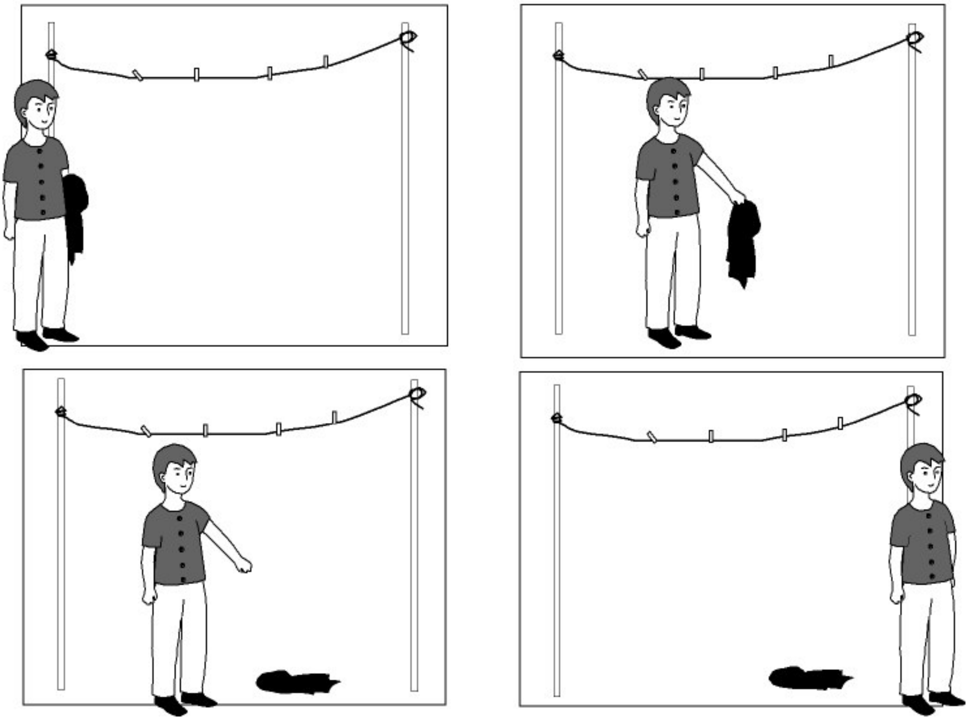
It should be noted that although truth value judgement tasks are generally designed such that the targetlike reading of the test sentence is falsified by the experimental context and the non-targetlike reading is verified, this is not possible here. As discussed in §2.1.2, the set of circumstances which verify the non-specific reading also verify the specific reading. For example, a context where a boy does not catch any fish from a group of three fish is consistent with the non-specific reading because it is true that the boy did not catch any fish, but it is also consistent with the specific reading of the indefinite because there is a fish which the boy did not catch (any one of the three) (see Musolino 1998 for relevant discussion).

There were five target items, using the verbs *vissen* 'to fish', *plukken* 'to pick', *stelen* 'to steal', *strijken* 'to iron' and *pakken* 'to get/take'.[4] These were interspersed with eight filler items in a pseudo-random order, which was used with all subjects. All the fillers contained a sentence with a scrambled order; four required a no-response and four a yes-response. Six were taken from Krämer's original experiment and two more of the same design were added to increase the ratio of fillers to test items. The fillers were used as controls and as distracters. They were designed such that a targetlike response did not depend on the subject having knowledge of scrambling, that is, the yes-fillers were always true and the no-fillers always false irrespective of whether the indefinite was interpreted within the scope of negation. An example of a yes-filler is given in (4).

In this scenario, there is one T-shirt, which the boy fails to wash. The filler sentence in (4) is thus true if the indefinite object is interpreted within the scope of negation ('It is not the case that the boy washed a T-shirt') or outside of the scope of negation ('There is a T-shirt which the boy did not wash'). Note, furthermore, that these fillers also function as a check on whether subjects can cope with true negatives. This is essential given that the test items are also true negatives and children are known to have difficulties with such sentences (De Villiers and Tager-Flusberg 1975; Wason 1965).[5]

[4] Due to time constraints (because the two (or three) tasks were carried out in a single session – see below), I decided to include one item less per condition than in Krämer's original experiment.

[5] Philip (2003:12) chooses to ignore Krämer's findings on scrambling across negation because, he claims, she did not control for the potentially confounding effect of children's difficulties with true negatives. As the discussion above indicates, however, this effect was controlled for in this experiment.



(4) Example scenario for 'yes-response' filler in truth value judgement task on scrambling across negation

Picture 1 Dit is een jongen. En dit is een T-shirt. Dat wil ie denk ik ophangen.
'This is a boy. And this is a T-shirt. I think he wants to hang it up.'

Picture 2 Maar, hier houdt ie een T-shirt vast…
'But here he's holding a T-shirt …'

Picture 3 …en hier laat ie het op de grond vallen.
'… and he he lets it fall on the floor.'

Picture 4 En nu gaat ie weer weg.
'And now he's leaving.'

Puppet:	De jongen	heeft	een T-shirt	niet	gewassen [target: acceptance]
	the boy	has	a T-shirt	not	washed

'The boy hasn't washed a T-shirt.'

As in Krämer's original experiment, the intonation used for all items was the most natural for the scrambled order, namely with the pitch rise for the VP starting on the negator *niet* 'not'. The indefinite determiner was always pronounced as *een* \ən\ rather than *één* 'one' (see §2.1.2 and §5.4.3.3 for relevant discussion).

6.3.1.2 Procedure

The experiment was carried out as follows. Subjects were told that they were going to listen to stories about some pictures in a book (a ring-binder containing colour pictures inserted into see-through plastic sleeves). They were then introduced to a silly puppet, *Karel de Kikker* 'Charles the Frog', who made a comment at the end of each story. Karel was allowed to listen to the stories but he was not allowed to look at the pictures. The subjects' task was therefore to use the information they saw in the pictures to say whether he was right or wrong. As in the production experiment, the puppet played a very active role with the younger children, but for the older children and adults, he was used to explain what the subjects should do and subsequently placed to one side. Specifically, the subjects were shown how the puppet was used with the younger children and once this was clear, the experimenter said that she would play the role of the puppet. For all subjects, the experimenter pointed to each picture in the story at the appropriate moment; she also clearly pointed to the final, unmanipulated object in the last picture several times, ensuring that the subjects had seen it.

As with the production task, the data collection procedure for the L2 subjects started with a brief explanation of the general purpose of the experiment, namely comparing younger and older second language learners. For the L2 children, who were usually excused from their regular lessons to take part, this helped to put them at ease. The fact that subjects could speak two languages was always commented on positively in order to emphasise the positive nature of the experiment. They were also told that it was not a test and that what they did had no repercussions for their schoolwork. Subjects were subsequently asked a few general questions regarding their background and language learning history. This allowed us to confirm that subjects fulfilled the criteria for participation (raised in a monolingual English environment, age of arrival after four years, etc.).

For all subjects, the experimental session started with a short warm-up designed to familiarise them with the truth value judgement task. This included three items similar in nature to the fillers but without any scrambling. Each sentence contained an intransitive verb. Two required a no-response and one a yes-response. Details are provided in Appendix D. When subjects failed to provide a targetlike response in the warm-up session, the experimenter helped them and explained the task (or 'the rules of the game' for the child subjects) once again. Only once it was clear that the subject understood how to complete the task did the experimenters proceed to the main experiment. Older L2ers and native adult controls were instructed to use their intuitions rather than trying to work out what the experiment was about and that if they so wished, more details regarding the focus of the experiment could be provided at the end of the session.

The truth value judgement task on negation was the third task to be carried out in the experimental session. It was preceded by the truth value judgement task on *twee keer* (see §6.4) and then the picture description task designed to collect proficiency data. This order of experiments was employed because the adult L2ers carried out a second (act-out) task on scrambling across *twee keer* and it was felt necessary that in order to minimise any potential influence of one task on the other, the two tasks involving *twee keer* should not directly follow each other. In addition, the picture description task served as a distracter/filler between the two truth value judgement tasks. In order to ensure comparability across subjects, the L1 children also completed the picture description task.[6]

Each subject was tested individually in a quiet room (at the subject's school or at Utrecht University or the University of Amsterdam) and each session was recorded using a Sony Hi-8 video recorder and a Tascam DA-P1 DAT recorder with a Crown PZM185 microphone. Child L1 subjects were rewarded for their participation with a small snack, the child L2 subjects with a small gift (such as stickers or a pen) and the adult subjects with a piece of cake. The L1 children and some of the L1 adults were tested by two native-speaker student assistants, who were trained in the experimental procedure. The L2 children and L2 adults and the remaining L1 adults were tested by one native-speaker student assistant and one near-native speaker (the present author). One experimenter told the stories and elicited the responses from subjects and the other played the role of the puppet. The session was conduced in Dutch as far as possible. If the L2 subjects experienced problems understanding, a short explanation was provided in English. However, those subjects who clearly did not understand the task were excluded. The experimental session, including the two truth value judgement tasks and the picture description task, lasted between 25 and 35 minutes. The youngest (L1 and L2) children who were tested were 7-years-old and they did not have any problems taking part in a session of this length. Adding the act-out task, the experimental session lasted around 45 minutes.

6.3.2 L1 acquisition

The purpose of carrying out the truth value judgement task with L1 children is twofold. First, given that the children in Krämer's negation experiment provided on average just 16% (25/156) targetlike responses, it is necessary to test older children in order to establish the age at which L1 children reach adultlike levels. Second, in order to have L1 data which are maximally comparable to the L2 data, it is essential to carry out with L1 children exactly

[6] The data collected during this task turned out to be vital in the design of the proficiency measure – see Chapter 4 for details.

the same experiment used with the L2ers.[7] The results are first analysed per age group (in §6.3.2.1) and subsequently, in terms of the individual children's response patterns (in §6.3.2.2).

6.3.2.1 Group results

Two subjects (L18#15C and L112#7C) were excluded because they failed on more than one filler.[8] Of the remaining fillers, 98.8% (742/751) were correct. Turning to the test items, the results for the six different age groups and the native adults are provided in Table 37.[9]

---

[7] The procedure used with the 13-year-old L1 children was, however, slightly different from the one outlined above. The stories were told in exactly the same way, but the subjects were tested as a group rather than individually. The pictures were presented on an overhead projector and subjects were asked to circle 'yes' or 'no' on an answer sheet in response to each item and, also, to motivate their responses. No conferring was allowed. Due to time constraints, these subjects were tested on four items only (instead of five). The reason this group was tested in this way was because after the first round of testing with 7-, 9- and 11-year-olds, it became clear that the oldest children did not always provide targetlike responses. In order to establish at what age the majority of L1 children were consistently targetlike, it was decided that a group of 13-year-olds should be tested as quickly as possible, and that if they were targetlike, a group of 12-year-olds and if necessary another group of 13-year-olds would be tested individually in order to complete the data set. In the end, the 12-year-olds turned out not to differ from these 13-year-olds. The data from the 13-year-olds are nevertheless included for the sake of completeness.

[8] The subject code should be read as follows: L1 denotes an L1 child, 8#15 gives the age group (8) followed by the number of the child in that group (15) and C indicates that this is the comprehension experiment.

[9] The production results in Chapter 5 were presented as the average percentage of targetlike responses per group because, as a result of the number of tokens differing per subject, it was this figure which was employed in the statistical tests which were applied there. In order to ensure maximal comparability with those data, the data here are also presented as average percentage of targetlike responses. Note, however, that because – in contrast to the production data – the number of tokens per subject in the comprehension experiment is identical (n=5), the (more conventional) percentage of targetlike responses per group could also be used as these two values are the same. For the sake of completeness, I include the number of tokens for each group as a whole in Table 37 here: 7-year-olds = 7/70; 8-year-olds = 15/75; 9-year-olds = 49/75; 11-year-olds = 43/75; 12-year-olds = 54/70; 13-year-olds = 77/100 and adults = 65/70.

Table 37. L1 children (comprehension): Group results for negation experiment

Subjects		Average percentage of targetlike (i.e. 'yes') responses	
Age group	**n**	**%**	**SD**
7-year-olds	14	10.0	25.7
8-year-olds	15	20.0	37.0
9-year-olds	15	65.3	41.7
11-year-olds	15	57.3	44.0
12-year-olds	14	77.1	29.6
13-year-olds	25	77.0	33.8
Adults	14	92.9	14.9

Table 37 shows that the youngest children regularly reject the specific interpretation for the indefinite object scrambled across negation. The motivation the children give for their answers indicates that they attribute a non-specific interpretation to the scrambled indefinite object NP. An example is provided in (5).

(5) Test sentence: Het meisje heeft een appel niet geplukt
the girl has an apple not picked
'The girl didn't pick a (certain) apple.'

Subject L17#5C: Nee, ze heeft twee appels geplukt
no she has two apples picked
'No, she picked two apples.'

When subjects provided a reason for their targetlike responses, as in (6), this indicated that the scrambled indefinite sentence was accepted for the right reasons, that is, because it was a true description of the specific reading for the indefinite depicted in the experimental scenario.

(6) Test sentence: Het meisje heeft een koekje niet gepakt
the girl has a cookie not taken
'The girl didn't take a (certain) cookie.'

Subject L19#15C: Ja, ze heeft twee koekjes gepakt
yes she has two cookies taken
'Yes, she took two cookies…

en één niet
and one not
…and not one (of them) .'

There is a significant difference between groups (ANOVA: df = 6, F = 16.483, p = .000). Post-hoc analyses were carried out in order to determine which groups differed from each other.[10] Significant differences (at the .05 level) are highlighted.

Table 38. L1 children (comprehension):
Results of post-hoc Games-Howell test for negation task

Comparison		**Mean difference %**	**Significance**
7-year-olds	8-year-olds	-10.0	p = .955
	9-year-olds	-55.3	p = .003
	11-year-olds	-47.3	p = .018
	12-year-olds	-67.1	p = .000
	13-year-olds	-67.0	p = .000
	adults	-82.9	p = .000
8-year-olds	9-year-olds	-45.3	p = .041
	11-year-olds	-37.3	p = .155
	12-year-olds	-57.1	p = .001
	13-year-olds	-57.0	p = .001
	adults	-72.9	p = .000
9-year-olds	11-year-olds	8.0	p = .995
	12-year-olds	-11.8	p = .946
	13-year-olds	-11.7	p = .938
	adults	-27.5	p = .256
11-year-olds	12-year-olds	-19.8	p = .705
	13-year-olds	-19.7	p = .675
	adults	-35.5	p = .100
12-year-olds	13-year-olds	1.4	p = 1.000
13-year-olds	adults	-15.9	p = .420

The 7-year-olds differ significantly from all groups except the 8-year-olds, and the 8-year-olds differ significantly from all groups except the 11-year-olds (and the 7-year-olds). The older groups (9 years and older) do not significantly differ from each other, nor from the adult controls.

6.3.2.2 Individual results

Following Krämer (2000), individual subjects were categorised according to the response pattern they produced. Subjects who (incorrectly) rejected all or all but one of the test items were classified as having a pattern of rejection. Those who (correctly) accepted all or all but one of the test items were categorised as having a pattern of acceptance, and all subjects in-between were assigned to a 'mixed' response pattern. The distribution of these response patterns across

---

[10] Given that the equal variances could not be assumed (Levene statistic = 3.082, p = .013), the Games-Howell post-hoc test was used. Repeated applications of the non-parametric Mann-Whitney test could also have been employed; this gives the same relative pattern of results as in Table 38.

the different age groups is presented in Figure 2. The exact numbers and individual data are provided in Appendix D.

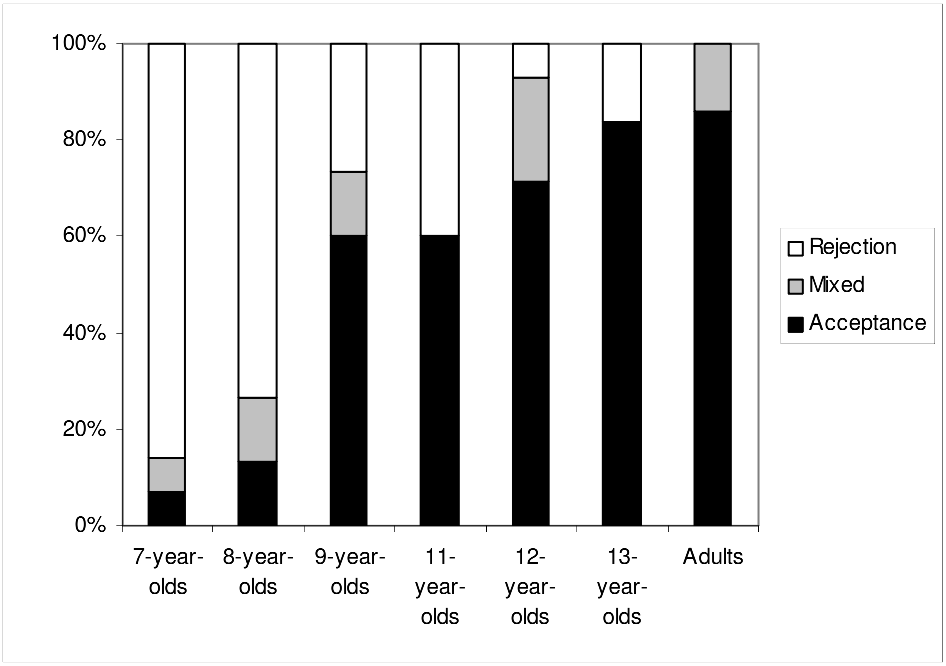


Figure 2. L1 children (comprehension): Distribution of individual response patterns on negation task

Most (12/14) of the adults accept all or all but one of the test items. The remaining two adult control subjects accept three out of the five items. The majority of the older children also have a pattern of acceptance. Most of the children in the younger two groups, however, have a rejection pattern. A clear developmental pattern is thus found, with a difference between children younger than nine years old, on the one hand, and children aged nine years and older, on the other. This tallies with the differences observed in Table 38.

6.3.2.3 Discussion

The results presented above show that 7- and 8-year-old L1 Dutch children generally fail to interpret an indefinite NP object scrambled across negation in a targetlike way. This is completely consistent with Krämer's (2000) findings, where 4- to 7-year-old Dutch children had an average acceptance rate of 16% (25/156). The data presented here show a leap towards targetlike behaviour at age 9. Although the older children have not quite reached targetlike levels of acceptance, there is no statistically significant difference between children aged nine and older and the adult controls.[11]

This gradual pattern of development is also consistent with the claim that the non-targetlike responses result from limited discourse integration. For example, in her study on the comprehension of definite articles, Karmiloff-Smith (1979) found a comparable pattern of development (cf. §3.3.1.1). The 3- to 5-year-olds in her study understood the anaphoric function of definite NPs in just over 30% of test items. Children became more targetlike with increasing age: 8-year-olds scored 51%, 9-year-olds 83% and 10-year-olds 90%. Although the proportion of targetlike responses is slightly higher than those observed for the equivalent age groups here, the general pattern is comparable. Regarding the possible cause of limited discourse integration in children, the observation that the 12- and 13-year-olds are not completely targetlike suggests that it is unlikely that processing problems will be responsible. What exactly might be responsible for the observed results is discussed in Chapter 7 (§7.3).

The results presented here are also, in principle, consistent with Philip (2003). A discussion of the results of the present study and Philip's approach will be postponed until §6.4.2.3, however.

6.3.3 L2 acquisition

The purpose of carrying out this task with L2 children and adults is twofold: (i) to consider whether L2 children are affected by limited discourse integration in the same way as has been claimed for L1 children and thus whether they differ from L2 adults, and (ii) to determine whether there are English-speaking L2ers who are able to overcome the poverty-of-the-stimulus to acquire the specific interpretation associated with scrambled indefinite objects. Exactly the same task was used with the L2 subjects as with the L1 subjects. The group results are presented first, in §6.3.3.1, followed by the individual results in §6.3.3.2.

Recall that the English equivalent of the Dutch test sentence in this task (e.g. 'The boy didn't catch a fish') is ambiguous, that is, the indefinite object could be interpreted specifically or non-specifically. What the L2ers must acquire is the fact that the Dutch sentence has only a specific reading. One consequence of the ambiguity in English is that if L2 subjects accept the specific reading of the scrambled indefinite object, it cannot automatically be assumed that they have targetlike knowledge of the relevant interpretive constraints, as such a response could potentially be the result of L1 transfer. The prediction is that when subjects accept the specific reading for the right reason (that is, because of knowledge of the target constraint), the proportion of subjects with targetlike responses should increase with proficiency. As the discussion in §3.3.1.3 highlighted, it is possible that the younger L2 children (at time of testing) might provide non-targetlike responses because they cannot

[11] This does not of course mean that these two groups are necessarily identical.

access the specific interpretation at all (including when interpreting equivalent sentences in their L1) rather than because of a non-targetlike interlanguage grammar.

6.3.3.1 Group results

Two children (C16C and C27C) were excluded because they failed on more than two fillers, indicating that they could not cope with true negatives. Virtually all (98.8% (561/568)) of the remaining fillers were answered correctly. As with the production experiment, the results will first be presented per proficiency group. Subsequently, the child L2 data are examined in terms of age at time of testing.

6.3.3.1.1 Proficiency level

As in Chapter 5, the child L2 and adult L2 groups were divided into three proficiency groups – low, mid and high – on the basis of their proficiency score (see fn 21 and 22 in Chapter 5 for more details). The analysis in this section will determine whether in their average percentage of targetlike responses, the various proficiency levels differ from native adults, whether they differ from each other within and across the two L2 groups, and whether there is evidence for development in targetlike knowledge of scrambling with increasing proficiency. An overview of the proficiency groups is given in Table 39.[12]

Table 39. L2 children and adults (comprehension): Overview of proficiency groups

	L2 children			L2 adults			Mann-Whitney tests comparing child/adult groups
	n	mean	SD	n	mean	SD	
Low	14	-1.20	0.39	18	-1.07	0.40	Z = -.866, p = .386
Mid	6	-0.02	0.16	8	0.12	0.31	Z = -1.162, p = .245
High	16	0.97	0.36	11	0.98	0.29	Z = -.444, p = .657

The results for the Mann-Whitney tests (two-tailed), given in the final column of Table 39, indicate that there is no significant difference between the child and adult groups in any of the three proficiency levels.[13] There are more low proficiency subjects within the L2 adult group than within the L2 child group; the distribution of the three proficiency levels across each of the two L2 groups is not significantly different, however (Chi-squared: $\chi^2 = 1.698$, df = 2, p = .428).

---

[12] This includes two L2 children and six L2 adults whose proficiency in Dutch was so low, they were unable to carry out the task. These subjects do not have a proficiency score but they are included in the low proficiency group in this and subsequent tables.

[13] The non-parametric Mann-Whitney test is used for data which do not conform to the prerequisites for a t-test. All the results using this test reported here are based on positive ranks.

Table 40 presents the average percentage of targetlike responses per proficiency group (cf. fn. 9 above on how results are presented).

Table 40. L2 children and adults (comprehension): Group results (organised by proficiency) on negation task. Average percentage of targetlike (i.e. yes) responses

	L2 children			**L2 adults**			**Mann-Whitney tests comparing child/adult groups**
	n	**%**	**SD**	**n**	**%**	**SD**	
Low	14	35.0	46.0	18	35.6	43.7	Z = -.047, p = .962
Mid	6	50.0	41.5	8	2.5	7.1	Z = -2.727, p = .006
High	16	26.3	39.8	11	40.0	45.6	Z = -.820, p = .412

The average rate of targetlike responses falls between 26.3% and 50.0% for all groups except the mid-proficiency L2 adults, who produce very few targetlike answers with an average of 2.5%. The L2 subjects motivate their non-targetlike responses in a similar fashion to the L1 children:

(7) a. Test sentence: De boef heeft een ketting niet gestolen
the thief has a necklace not stolen
'The thief didn't steal a (certain) necklace.'

Subject C1C: Nee, de boef heeft WEL een ketting gestolen
no the thief has indeed a necklace stolen
'No, the thief DID steal a necklace.'

b. Test sentence: De jongen heeft een vis niet gevangen
the boy has a fish not caught
'The boy didn't catch a (certain) fish.'

Subject A29C: Nee, hij heeft twee vissen gevangen
no he has two fish caught
'No, he caught two fish.'

In response to the scenario where a girl took two out of three cookies, one subject (A34C), who responded in English rather than Dutch, even explicitly mentioned the non-specific interpretation in her rejection of the test sentence:

(8) No, because she took two of them and you said she took none of them.

Note, however, that it is unlikely that these non-targetlike responses resulted from subjects' failure to notice the last, unmanipulated object, as this object was regularly commented upon:

(9) a. Test sentence: De boef heeft een ketting niet gestolen
the thief has a necklace not stolen
'The thief didn't steal a (certain) necklace.'

	Subject C25C:	Nee,	maar	hij	heeft	er	wel	ééntje	achtergelaten
		no	but	he	has		indeed	one	behind-left

'No, but he did leave one of them behind.'

Ik	zou	die	ook	achterlaten
I	would	that	also	behind-leave

'I would leave that behind, too …

want	hij	is	niet	mooi
because	he	is	not	beautiful

… because it's not beautiful.'

b.	Test sentence:	De boef	heeft	een ketting	niet	gestolen
		the thief	has	a necklace	not	stolen

'The thief didn't steal a (certain) necklace.'

Subject C9C:	Nee,	hij	heeft	'm	wel,	twee	gestolen
	no	he	has	it	indeed	two	stolen

'No, he has *stolen it, two.'*

Een boef	wil	stelen	altijd	diamanten	dingen
a thief	wants	steal	always	diamond	things

'A thief always wants to steal diamond things.'

Maar	ze	laten	dingen	wel	staan	soms
but	they	leave	things	indeed	stand	sometimes

'But they do leave things behind sometimes.'

On several occasions, the form of the NP clearly influenced subjects' responses. After hearing the test sentence with the scrambled indefinite object realised as the unstressed indefinite determiner *een* 'a', pronounced as \ən\, certain subjects enquired as to whether the experimenter had said *een* or *één*. At this point, the experimenter simply repeated the test sentence with exactly the same pronunciation. As the following sequences demonstrate, subjects were often unwilling to accept the specific interpretation for the scrambled indefinite unless it was realised as *één*. Subject A30C in (10)-c appears to have heard the indefinite as *één* and consequently (correctly) accepts it.

(10)	a.	Test sentence:	Het meisje	heeft	een koekje	niet	gepakt
			the girl	has	a cookie	not	taken

The girl didn't take a (certain) cookie.'

Subject A22C:	Een koekje	of	één koekje?
	A cookie	or	one cookie

'A cookie or one cookie?'

The experimenter then repeats the test sentence as above and the subject rejects it.

b. Test sentence: De vrouw heeft een bloes niet gestreken
the woman has a blouse not ironed
The woman didn't iron a (certain) blouse.'

Subject A28C: EEN \ən\?
a
'A?'

The experimenter repeats the test sentence as above.

Klopt niet. Ze heeft twee bloezen gestreken
is-right not she has two blouses ironed
'That's not right. She ironed two blouses.'

c. Test sentence: De vrouw heeft een bloes niet gestreken
the woman has a blouse not ironed
The woman didn't iron a (certain) blouse.'

Subject A30C: Ze heeft WEL een bloes gestreken,
she has indeed a blouse ironed
'She DID iron a blouse…

maar ze heeft wel TWEE bloezen gestreken
but she has indeed two blouses ironed
… but she DID iron TWO blouses.'

The experimenter repeats the test sentence as above.

ÉÉN? Goed.
one good
'One? Good.'

d. Test sentence: De boef heeft een ketting niet gestolen
the thief has a necklace not stolen
'The thief didn't steal a (certain) necklace.'

Subject C15C: Één ketting niet gestolen of ….?
one necklace not stolen or
'One necklace not stolen or…?'

The subject pauses briefly and then decides that the test sentence is correct.

These responses suggest that subjects associate the specific meaning with the form (*één* vs. *een*) of the NP rather than its position (scrambled vs. non-scrambled). One subject even states this explicitly, as in (11).

(11) a. Test sentence: Het meisje heeft een koekje niet gepakt
the girl has a cookie not taken
The girl didn't take a (certain) cookie.'

Subject A17C:	Ik	zou	zeggen	dat	ze	heeft	wel	
	I	would	say	that	she	has	indeed	
	'I would say that she has …							
	een koekje		gepakt,	maar	als	je	'één'	zegt
	a cookie		taken	but	if	you	one	say
	… *taken a cookie, but if you say 'one'* …							
	dan	is	het	anders				
	then	is	it	different				
	… *then it's different.'*							

When subjects accept the test sentences, their comments, exemplified in (12), indicate that this is for the right reasons, that is, they assign a specific interpretation to the scrambled utterance.[14]

(12)	Test sentence:	Het meisje	heeft	een koekje	niet	gepakt	
		the girl	has	a cookie	not	taken	
		'The girl has taken a (certain) cookie.'					
a.	Subject C29C:	Ja.	Zij	laat	één	op tafel	
		yes	she	leaves	one	on table	
		'Yes, she leaves one on the table.'					
b.	Subject A36C:	Ja,	want	ze	heeft	twee koekjes	gepakt
		yes	because	she	has	two cookies	taken
		'Yes, because she has taken two cookies…					
		en	ze	heeft	één	op de tafel	gelaten
		and	she	has	one	on the table	left
		… *and she left one on the table.'*					

The results for the Mann-Whitney tests given in Table 40 indicate that the child L2 and adult L2 low proficiency groups do not significantly differ from each other and neither do the two high proficiency groups. There is, however, a significant difference between the mid proficiency groups. This is clearly due to the very low proportion of targetlike responses in the adult L2 mid proficiency group. Comparing the percentage of targetlike responses for the three proficiency levels and the native adults, a significant difference is observed for both the L2 children (ANOVA: df = 3, F = 9.368, p = .000) and the L2 adults (ANOVA: df = 3, F = 13.364, p = .000). Post-hoc analyses were carried out to determine within the child and adult groups, which proficiency groups differed from each other and/or from the native adults. The results of

[14] Participants in a truth value judgement task rarely motivate yes-responses spontaneously and if experimenters ask them to do so, it can sound somewhat infelicitous. It can also lead the subject to doubt his or her answer. For this reason, the subjects here were only asked to provide motivation for their targetlike yes-response on the final item in the task. In this way, our request could not have any adverse effect on subsequent items.

these analyses are presented in Table 23, where significant differences are highlighted.[15]

Table 41. L2 children and adults (comprehension): Results of post-hoc Games-Howell test for negation task

Comparison	Mean difference %	Significance
Child L2 Low vs. Child L2 Mid	-15.2	p = .896
Child L2 Low vs. Child L2 High	8.7	p = .952
Child L2 Mid vs. Child L2 High	24.2	p = .637
Child L2 Low vs. Adult natives	-58.8	p = .005
Child L2 Mid vs. Adult natives	-43.1	p = .171
Child L2 High vs. Adult natives	-67.0	p = .000
Adult L2 Low vs. Adult L2 Mid	33.6	p = .027
Adult L2 Low vs. Adult L2 High	-3.8	p = .994
Adult L2 Mid vs. Adult L2 High	-38.7	p = .087
Adult L2 Low vs. Adult natives	-57.1	p = .000
Adult L2 Mid vs. Adult natives	-90.1	p = .000
Adult L2 High vs. Adult natives	-53.2	p = .015

Within the child L2 group, none of the proficiency levels significantly differ from each other, but the low and high groups differ significantly from the native adults. The low proficiency L2 adults differ significantly from the mid proficiency L2 adults, and there is a statistically significant difference between each of the three L2 adult proficiency levels and the native adults. The difference between the mid proficiency L2 adults and the high proficiency L2 adults is approaching significance.

There was no correlation between targetlike responses and length of exposure (L2 children: r = .057, p = .748; L2 adults: r = .159, p = .346), proficiency score (L2 children: r = -.042, p = .819; L2 adults: r = -.089, p = .634) or age at time of testing (L2 children: r = .057, p = .748; L2 adults: r = .159, p = .346). In the following section, the variable age at time of testing will be examined in more detail.

6.3.3.1.2 Age at time of testing

The results presented in §6.3.2 indicated that there is an effect of chronological age for L1 children, that is, the youngest children, the 7- and 8-year-olds, were found to differ significantly from the older children and the native adults. In

[15] Games-Howell post-hoc tests were used because the assumption of equal variances between groups was not fulfilled for either L2 group (Levene statistic = 7.326, p = .000 for the L2 children and 21.604, p = .000 for the L2 adults). The assumption of equal variances was also not met for the child L2 group in the *twee keer* scrambling condition (Levene statistic = 7.764, p = .000) and for the adult L2 group in the *twee keer* no scrambling condition (Levene statistic = 4.137, p = .011).

order to determine whether such an age-effect is also at play in the child L2 data, the child L2 group results are now analysed according to chronological age, that is, on the basis of age at time of testing. Note that when the data were collected, the necessity of such an analysis was not apparent (because the L1 data had not yet been collected). Consequently, the L2 children's age at testing was not used as a criterion for selection and, as a result, the number of subjects is not evenly distributed across each age group. The 13- and 14-year-olds form the oldest group; unfortunately, there were no 12-year-olds.[16] The average percentage of targetlike responses per age group is given in Table 42. Table 42 also gives the mean and the range of the proficiency scores per age group. There is a significant difference between the different age groups in terms of proficiency score (ANOVA: F = 3.193, df = 5, p = .021), but post-hoc Bonferroni tests indicate that it is only the difference between the 7-year-olds and the 13/14-year-olds which is approaching significance (mean difference = -1.64, p = .095).

Table 42. L2 children (comprehension): Group results (organised by at time of testing) on negation task. Average percentage of targetlike (i.e. yes) responses

Subjects		**Targetlike responses**		**Proficiency score[a]**			**Proficiency group(s)**
Age group	**n**	**mean %**	**SD**	**mean**	**range**	**SD**	
7-year-olds	3	0	0	-0.62	-1.98 to 0.74	1.20	2 low, 1 high
8-year-olds	10	38.0	49.4	-0.59	-1.52 to 1.41	.99	7 low, 1 mid, 2 high
9-year-olds	3	13.3	23.1	0.36	-0.51 to 0.70	.61	1 low, 1 mid, 1 high
10-year-olds	5	48.0	50.2	0.57	-0.66 to 1.49	.93	1 low, 1 mid, 3 high
11-year-olds	5	48.0	50.2	0.50	-0.15 to 1.13	.50	2 mid, 3 high
13/14-year-olds	8	30.0	33.8	0.62	-1.36 to 1.71	.90	1 low, 1 mid, 6 high

[a] The proficiency level for one 7-year-old and one 8-year-old was so low that they were unable to complete the task.

The data in Table 42 suggest that the older children (ten years and above) generally produce more targetlike responses than the younger children (nine years and younger), but no clear developmental pattern can be discerned. The standard deviations indicate that there is considerable variation in most groups. This is not unexpected given that virtually all six age groups contain children at different levels of proficiency. An analysis of variance comparing these groups and the native adults is statistically significant (ANOVA: df = 6, F = 5.294, p = .000). However, post-hoc Games-Howell tests indicate that some of the child L2 age groups differ from the native speakers, rather than from each other. The 7- and 13/14-year-olds differ significantly from the native adults

[16] This was not for want of trying: all international schools in The Netherlands were contacted in order to find suitable subjects. Unfortunately, however, the only (two) 12-year-olds located were first exposed to Dutch before the age of four (see Appendix D for details).

and the difference between this latter group and the 8- and 9-year-old L2 children is also approaching significance.

Table 43. L2 children (comprehension):
Results of post-hoc Games-Howell test on negation task

Comparison		**Mean difference %**	**Significance**
7-year-olds	8-year-olds	-38.0	p = .285
	9-year-olds	-13.3	p = .917
	10-year-olds	-48.0	p = .465
	11-year-olds	-48.0	p = .465
	13/14-year-olds	-30.0	p = .281
	native adults	-92.9	p = .000
8-year-olds	9-year-olds	-24.7	p = .876
	10-year-olds	-10.0	p = 1.000
	11-year-olds	-10.0	p = 1.000
	13/14-year-olds	-9.0	p = 1.000
	native adults	-54.9	p = .068
9-year-olds	10-year-olds	34.7	p = .819
	11-year-olds	34.7	p = .819
	13/14-year-olds	-16.7	p = .952
	native adults	-79.5	p = .087
10-year-olds	11-year-olds	0	p = 1.000
	13/14-year-olds	-18.0	p = .987
	native adults	-44.9	p = .528
11-year-olds	13/14-year-olds	18.0	p = .987
	native adults	-44.9	p = .528
13/14-year-olds	native adults	-62.9	p = .010

Before turning to the individual results, let us consider the L2 adult group in terms of age at time of testing. Recall that membership of the L2 adult group was defined in terms of age at first exposure, that is, at age eight or older. The age at time of testing in this group ranges from 9;8 to 50 years (mean = 24;0; SD = 13;4). This means that some of the L2 adults may be young enough for their responses on the scrambling task(s) to be affected by limited discourse integration. This could potentially distort the results of the L2 adult group. For this reason, the L2 adult data were split into younger L2 adults (age at time of testing: mean = 11;11, range = 9;8-14;11, SD = 1;7; n=18) and older L2 adults (age at time of testing: mean = 35;6, range = 22;2-50; SD = 8;2; n=19) and these two groups were compared with each other and the original adult L2 group.[17] The results are given in Table 44.

[17] For the older adult L2 group, the length of exposure ranged from 0;4 to 25;0 (mean = 6;11, SD = 6;4), age at first exposure from 21;10 to 43;0 (mean = 28;4, SD = 6;0) and proficiency score from –1.50 to 1.37 (mean = 0.14, SD = 0.88). For the younger adult L2 group, the length of exposure ranged from 0;2 to 5;2 (mean = 1;7, SD = 1;2), age at first exposure from 8;2 to

Table 44. Younger and older L2 adults (comprehension): Group results (organised by proficiency) on negation task. Average percentage of targetlike (i.e. yes) responses

	L2 adults (younger subjects only)			L2 adults (older subjects only)			L2 adults (original)		
	n	**%**	**SD**	**n**	**%**	**SD**	**n**	**%**	**SD**
Low	12	31.7	47.1	6	43.3	38.8	18	35.6	43.7
Mid	3	0	0	5	4.0	8.9	8	2.5	7.1
High	3	33.3	57.8	8	42.5	44.6	11	40.0	45.6

A brief glance at Table 44 indicates that in terms of percentage of targetlike responses per proficiency level, there is little difference between the younger and older adult L2 groups. However, in comparison with the original adult L2 group, there are significant differences between the three proficiency levels and the native adults for both the older L2 adults (ANOVA: $F = 13.626$, $df = 3$, $p = .000$) and the younger L2 adults (ANOVA: $F = 10.200$, $df = 3$, $p = .000$). The post-hoc analyses, given in Table 45, indicate that in the younger adult L2 group, it is the low and mid proficiency levels which differ significantly from the natives, and in the older adult L2 group, it is the mid and high proficiency levels which each differ significantly from the natives, with the contrast between the low proficiency older L2 adults and the natives also approaching significance. This is slightly different from the post-hoc analyses for the original adult L2 group, given in Table 23, where the (original) low proficiency adults differed from both the natives and the mid proficiency adults. That this within-group difference is now absent is perhaps unsurprising given that the number of subjects per group is considerably reduced in the analysis presented in Table 45.

12;5 (mean = 10;3, SD = 1;8) and proficiency score from –1.89 to 1.54 (mean = 0.28, SD = 1.05).

Table 45. Younger and older L2 adults (comprehension): Results of Games Howell post-hoc test on negation task

Comparison	**Mean difference %**	**Significance**
Younger Adult Low vs. Younger Adult Mid	31.7	p = .150
Younger Adult Low vs. Younger Adult High	-1.7	p = 1.000
Younger Adult Mid vs. Younger Adult High	-33.3	p = .768
Younger Adult Low vs. Natives	-61.2	p = .004
Younger Adult Mid vs. Natives	-92.9	p = .000
Younger Adult High vs. Natives	-59.5	p = .467
Older Adult Low vs. Older Adult Mid	39.3	p = .182
Older Adult Low vs. Older Adult High	0.8	p = 1.000
Older Adult Mid vs. Older Adult High	-38.5	p = .163
Older Adult Low vs. Natives	-49.5	p = .088
Older Adult Mid vs. Natives	-88.9	p = .000
Older Adult High vs. Natives	-50.4	p = .059

6.3.3.2 Individual results

We now examine the results for individual subjects. As with the group data, the individual data are first categorised according to proficiency level and subsequently, according to age at time of testing.

6.3.3.2.1 Proficiency level

As for the L1 children, each L2 subject was categorised according to whether s/he accepted or rejected all or all but one of the five test items. Subjects who accepted/rejected two or three items were classified as mixed. The results, organised per proficiency level, are presented in Figure 3 for the L2 children and Figure 4 for the L2 adults.[18] The exact numbers are given in Appendix D.

[18] As was the case for the individual production data presented in Chapter 5, the number of subjects per category is too small to carry out, for example, a Chi-squared test to determine whether the distribution of response patterns differs per proficiency level.

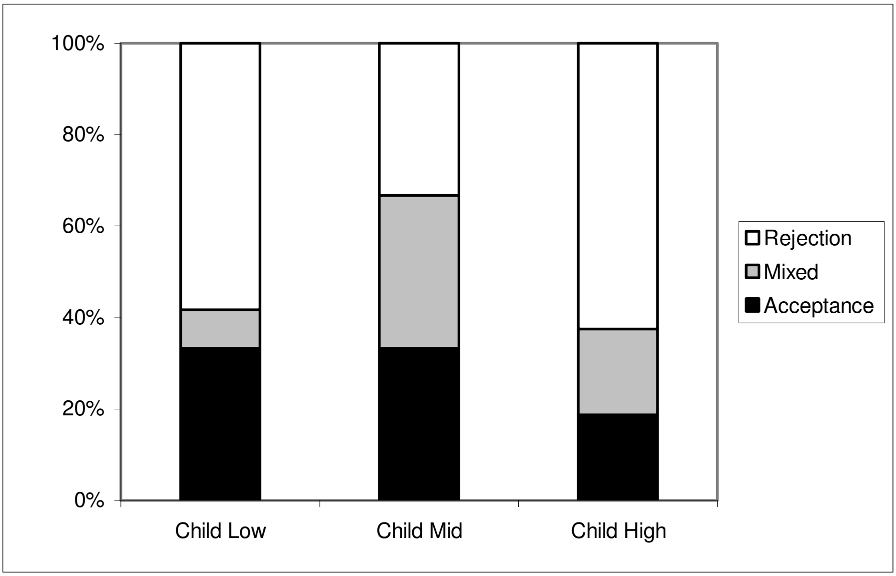


Figure 3. L2 children (comprehension): Distribution of individual response patterns (per proficiency group) on negation task

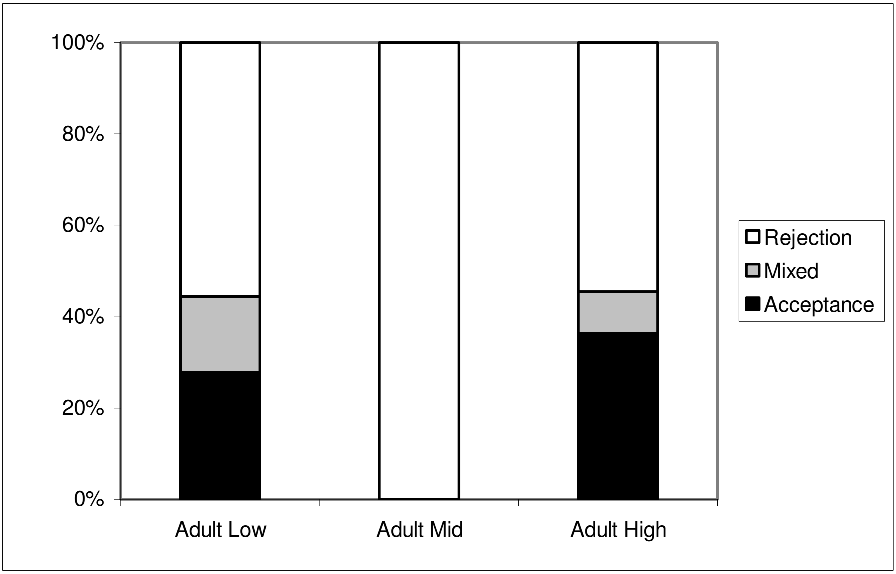


Figure 4. L2 adults (comprehension): Distribution of individual response patterns (per proficiency group) on negation task

With the exception of the mid proficiency L2 children, who are distributed evenly across the three response patterns, in each child/adult L2 proficiency group, the majority of subjects have a (non-targetlike) pattern of rejection. For the child L2 subjects, around one third of the low and mid groups and one fifth of the high group have a targetlike pattern of acceptance. Amongst the L2 adults, around one third of the low and high proficiency subjects have a pattern of acceptance. None of the mid proficiency L2 adults consistently accept the test items.

6.3.3.2.2 Age at time of testing

The individual child L2 data analysed according to time of testing are given in Figure 5. A non-targetlike pattern of rejection is observed in the majority of children in the youngest three age groups and in half of those in the oldest age group. In the 8-, 10- and 11-year-old groups, just under half of subjects have the targetlike pattern of acceptance. In the oldest group, there is just one subject with this pattern and three with a mixed pattern, that is, three subjects who sometimes correctly accept the specific interpretation for the indefinite object scrambled across negation.

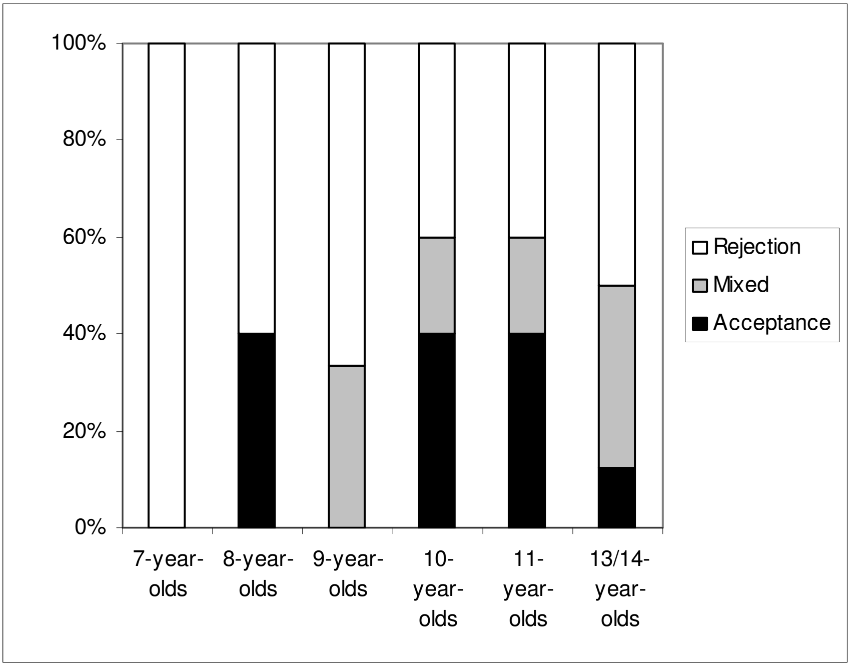


Figure 5. L2 children (comprehension): Distribution of individual response patterns (per age at time of testing) on negation task

The individual data for each age group are listed in Appendix D (organised by proficiency). The different number of subjects and the range of proficiency scores in each group render a cross-group comparison difficult. Although the children in the oldest group who have the targetlike pattern of acceptance and one of the two children with a mixed pattern have the highest proficiency scores in that group (1.09 and 1.71, respectively), there are four 8-year-olds with the target pattern who have much lower proficiency scores (one child could not complete the proficiency task and the scores for the other three range from –1.07 to –0.20). The children with an acceptance pattern in the other groups are distributed across the range of proficiency scores.

6.3.3.3 Discussion

The data presented here show that there are relatively few L2 subjects who provide targetlike responses on the negation task. This discussion considers several reasons for why this should be case. It addresses the role of L1 transfer, the lack of correlation between targetlike responses and proficiency, and the form of the object NP. Finally, the data are used to evaluate the predictions outlined in §6.1 on the role of discourse/pragmatic factors and the differences between the three learner groups.

There are both L2 children and L2 adults who consistently accept the specific interpretation for an indefinite object scrambled across negation and their comments during the task indicate that they do so for the right reasons. These targetlike subjects form just one quarter (25.4% (18/71)) of the total number of subjects, however, and within the child L2 and adult L2 groups, they are spread across the different proficiency levels. It is possible that for some of these subjects, the acceptance of the specific reading results from L1 transfer rather than targetlike knowledge. Without actually testing each individual subject in English as well, it is unfortunately not possible to distinguish between the two. What is clear, however, is that the predicted gradual increase in targetlike responses with increasing proficiency was not observed for either the child L2 or the adult L2 group.

Turning to proficiency, then, for both the L2 children and the L2 adults, no discernible developmental pattern is found, that is, it is not the case that the high proficiency groups are more targetlike than the mid and low proficiency groups, etc. Removing the younger subjects from the L2 adult group does not alter this finding. The absence of a positive relationship between targetlike comprehension of scrambled indefinite objects and proficiency could indicate that there is in fact no relationship between these two variables. It could also be the effect of the proficiency measure being based on production data only. In language development, it has often been observed that comprehension is more targetlike than production, in which case, the production task used here could underestimate subjects' actual proficiency level, at least for some subjects. This certainly appeared to be the case for several of the younger adult L2 subjects (A6C, A7C, A8C, A9C and A11C). Interaction with these subjects during the experimental session indicated that their understanding of Dutch was more than elementary, yet they were unable – and in certain cases, perhaps unwilling – to speak the language.[19] Two of these children produced targetlike responses some or all of

[19] It was my impression that the fact that these subjects were all in their early teens may also have contributed to their inability/unwillingness to speak Dutch; they appeared to be self-conscious about their accent, making errors, etc.

the time; if the proficiency task had included comprehension, they might have been able to complete it and their proficiency scores would have tallied better with their scrambling responses.

Several subjects commented upon the indefinite determiner, that is, whether it was realised as *een* or *één*. These comments were almost exclusively produced by the L2 adults, some of whom appeared to base their non-targetlike rejection of the test sentence on the fact that the element at issue was the indefinite determiner *een* rather than the cardinal/stressed indefinite *één*. The fact that there was only one child L2 subject who remarked on the indefinite determiner (see (10)-d) and that this subject actually accepted the test sentence does not of course necessarily mean that L2 children are not affected by this difference. In proficiency terms, the only two groups of L2ers who are *not* significantly different from the native controls are the child L2 mid proficiency level and the younger adult L2 high proficiency level. When subjects incorrectly reject the specific reading of the scrambled indefinite, they motivate their responses in a similar fashion to the L1 children. Note that the source of these non-targetlike responses could be the same as is claimed for L1 acquisition, that is, limited discourse integration, or it could be the result of L1 transfer.

Let us now turn to the predictions made on the basis of Krämer (2000). In §6.1, it was proposed that if the English-speaking L2 children have limited discourse integration, a correlation between age at time of testing and number of targetlike scrambling responses would be expected, irrespective of proficiency level. The observation that 7-, 8- and 9-year-olds were almost significantly different from the native adults suggests that age at time of testing could be a relevant factor. However, the fact that the 13-year-olds are significantly different from the native adults undermines this, especially as the 13-year-olds are generally in the high proficiency group. There was no correlation between targetlike responses and age at time of testing, nor between targetlike responses and proficiency. This latter finding suggests that L1 transfer is not the sole relevant factor either. It was suggested that if both limited discourse integration and L1 transfer were at play in child L2 acquisition, only the older, high proficiency child L2ers would have targetlike responses. This prediction is difficult to test, however, because the range of proficiency scores within the oldest child L2 group is rather limited, as are the number of subjects and the range of proficiency scores in the other child L2 groups. The observation that there are several high proficiency subjects in this oldest group who are non-targetlike speaks against it, nevertheless. If we thus take the existence of an age-effect in L1 acquisition to result from limited discourse integration, the lack of such an effect in the child L2 data discussed here suggests that limited discourse integration does not play a role in the child L2 comprehension of scrambled indefinite objects. Although there are some age-related differences in the child L2 data (for example the youngest groups are significantly different from the native adults), on the whole, in terms of age-related effects, these data contrast with those of the L1 children presented in §6.3.2 above. This suggests that prediction ❹, which stated that L2 children within the same age range as L1 children with limited discourse integration should behave similarly to L1 children in their interpretation of scrambled indefinite objects, is not borne out.

Let us now consider prediction ❷. This was split into two parts. The first part stated that in their interpretation of scrambled indefinite objects, L2 children within the same age range as L1 children with limited discourse integration would differ from L2 adults. The second part predicted that L2 children outside of this age range should pattern similarly to L2 adults. In order to evaluate the extent to which these two sub-predictions are borne out, it is first necessary to determine the age range in which (at least some) L1 children can be said to have limited discourse integration. I will assume (for the moment at least) that this is the age range in which L1 children do not provide adultlike levels of targetlike responses on the negation task. Given that the 13-year-old L1 children do not yet (as a group) provide adultlike levels of acceptance, it is therefore assumed that all the L2 children, who are between seven and 14 years old, are broadly speaking within the same age range as the L1 children. Consequently, the second part of prediction ❷ cannot be evaluated, because there are no L2 children outside of this range. Prediction ❷ can therefore be reformulated to simply state that all the L2 children in this experiment are expected to differ from the L2 adults. Such a difference between the two groups was expected to manifest itself at the high proficiency level; the high proficiency L2 children were expected to be significantly less targetlike than the high proficiency L2 adults. This prediction is not borne out, however. Although there is a difference in the average percentage of targetlike responses (40% for the adults and 26.3% for the children, see Table 40), both the high proficiency children and the high proficiency adults differ significantly from the native adults and there is no effect of proficiency in either, that is, unlike the production data on scrambling across negation, targetlike comprehension of scrambling across negation does not appear to increase as a function of proficiency.

6.3.4 Summary and conclusion

This section reported data from a truth value judgement task which tested whether L1 children, L2 children and L2 adults accept the specific interpretation for indefinite object NPs scrambled across negation. In accordance with previous findings, the development of targetlike responses in L1 children was gradual and – in comparison with many other aspects of language – rather late. The youngest child L1 groups, aged seven and 8, were significantly different from both older children and adults. It was only at age nine that L1 children were no longer significantly different from adults, but even at age 13, the rate of targetlike responses was not at the adultlike level. These results are in principle compatible with the claim that non-targetlike comprehension of scrambled indefinite objects is related to limited discourse integration.

The child L2 data were vaguely suggestive of a similar age-effect but on the whole, such an effect was absent. Although there were some child L2 and adult L2 subjects who consistently provided targetlike responses, they constitute a small proportion of the total number of subjects and, crucially, they were not necessarily those with the highest proficiency. In other words, targetlike interpretation of scrambling was not related to proficiency in either of the L2 groups. It was suggested that this might be due to (i) the nature of the L2 proficiency task which was based on production data only, and (ii) L1 transfer, which might manifest itself in a preference for the indefinite being realised as *één* rather than *een*. I return to the issue of L1 transfer in §6.6.1.3, where the results of the two truth value judgement tasks are compared.

The following section presents data from a comparable task on scrambling across *twee keer*. There, we will see that the results are quite different from those presented here.

6.4 Task II: Truth value judgement task with *twee keer*

The purpose of this experiment is to determine whether language learners assign a targetlike specific interpretation to an indefinite object NP scrambled across the adverb *twee keer* 'twice', as in (13).

(13) a. Het meisje heeft een aap twee keer gekieteld
the girl has a monkey two times tickled
'The girl tickled a (certain) monkey twice.'

b. Het meisje heeft twee keer een aap gekieteld
the girl has two times a monkey tickled
'The girl tickled a(ny) monkey twice.'

In the sentence in (13)-a, the scrambled object receives a specific interpretation, that is, it can only mean that there is a sinlge monkey which the girl tickled twice. In the non-scrambled sentence in (13)-b, however, the girl can tickle one and the same monkey or two different monkeys, as this sentence means that there are two monkey-tickling events.

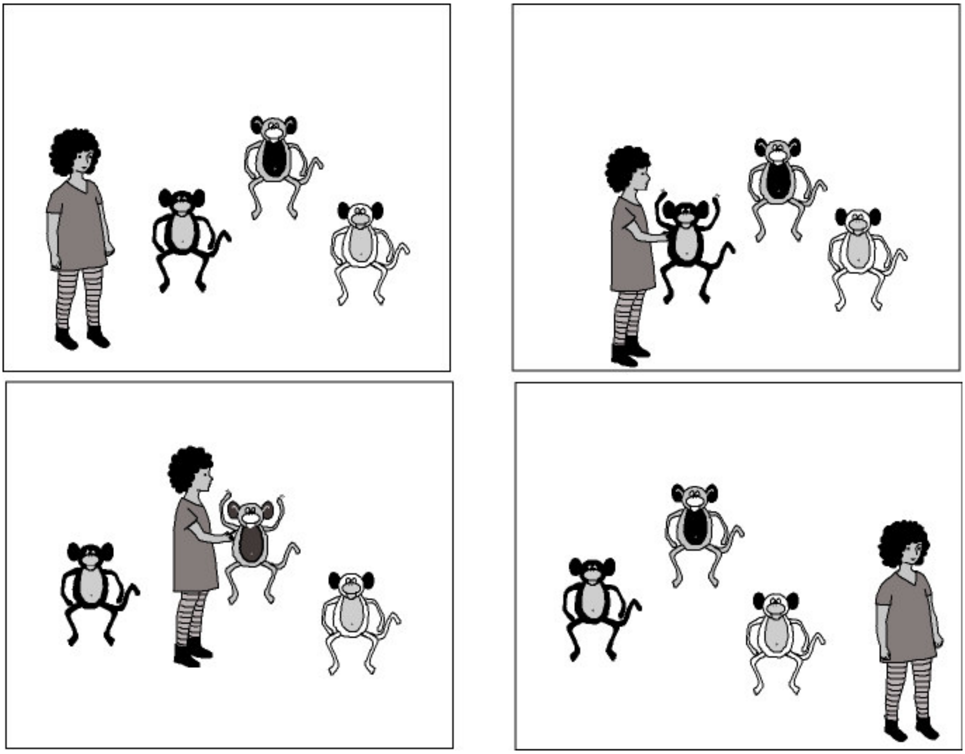
6.4.1 Method

A truth value judgement task was developed along the lines of Krämer (2000) and Philip (2003). Note that in contrast to the negation task, the *twee keer* task follows the usual set-up of a truth value judgement task, where the test sentence is falsified by the experimental context.

6.4.1.1 Materials

Subjects were presented with a series of four pictures which told a story about a girl or a boy. There were also three objects, differentiated from each other by colour. In the course of the story, the girl or boy carried out an action on two of the three objects. An example is given in (14). The story should be read from top left to top right, then from bottom left to bottom right.

Here, the subject is introduced to the protagonist and three different-coloured monkeys. The difference between the three objects is stated explicitly as this is a crucial feature of the story. (The pictures were in colour in the actual experiment.) In the second picture, the girl tickles one of the monkeys, the black one, and in the third picture, she tickles another monkey, the grey one. As a description of this context, the scrambled sentence in (14)-a above is incorrect: this sentence states that there is a monkey which the girl tickled twice and this is not the case because she tickled two different monkeys. The non-scrambled sentence in (14)-b is correct, however: it is the case that the girl tickled a monkey twice, that is, that there were two monkey-tickling events.



(14) Example scenario for experimental item in truth value judgement task on scrambling across *twee keer*

Picture 1 Dit is een meisje en dit zijn apen.
Een zwarte, een grijze en een witte. Die wil ze denk ik kietelen.
'This is a girl and these are monkeys.
A black one, a grey one and a white one. I think she wants to tickle them'

Picture 2 Hier kietelt ze een aap.
'Here, she's tickling a monkey'

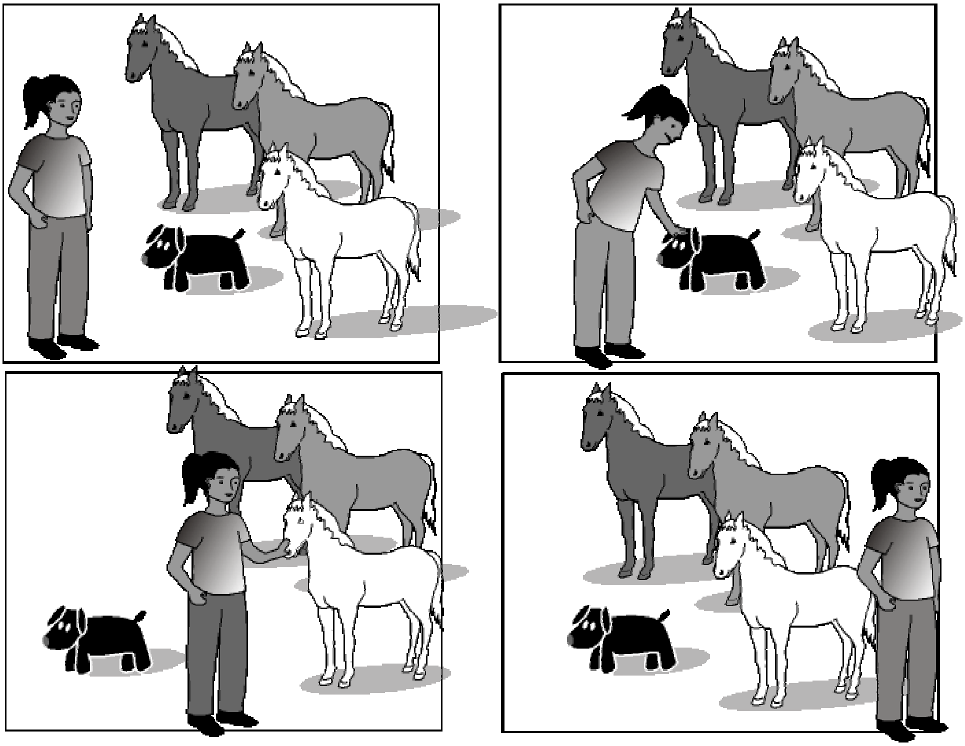
Picture 3 En hier kietelt ze een aap.
'And here she's tickling a monkey.'

Picture 4 En nu gaat ze weer weg.
'And now she's leaving.'

Puppet: a. Het meisje heeft een aap twee keer gekieteld [target: rejection]
the girl has a monkey two times tickled
'The girl tickled a (certain) monkey twice.'

b. Het meisje heeft twee keer een aap gekieteld [target: acceptance]
the girl has two times a monkey tickled
The girl tickled a(ny) monkey twice.'

There were five target items in each of the two (scrambled and non-scrambled) conditions. The following verbs were selected because they allow for iterative actions without a change-of-state: *kietelen* 'to tickle', *gooien* 'to throw', *aaien* 'to stroke/pet', *knijpen* 'to pinch' and *zoenen* 'to kiss'. The test items were interspersed with 16 fillers. These all contained the adverb *twee keer*; one half contained a sentence with a scrambled order and the other half a non-scrambled order. The targetlike response for the scrambled order in the test items is 'no'; consequently, in order to balance the number of 'yes' and 'no' responses, three of the eight scrambled fillers required a no-response and five a yes-response. Likewise, in the non-scrambled condition, where 'yes' was the targetlike response for the test items, three of the fillers required a yes-response and five a no-response. The fillers were used as controls and as distracters. They were designed such that a targetlike response did not depend on the subject having knowledge of scrambling, that is, the yes-fillers were always true and the no-fillers always false irrespective of whether the indefinite was interpreted within the scope of the adverb. The yes-fillers used a definite NP object and the no-fillers an indefinite object. Examples of each – with both the scrambled and non-scrambled orders – are given in (15) and (16), respectively.



(15) Example scenario for 'no-response' filler item in truth value judgement task on scrambling across *twee keer*

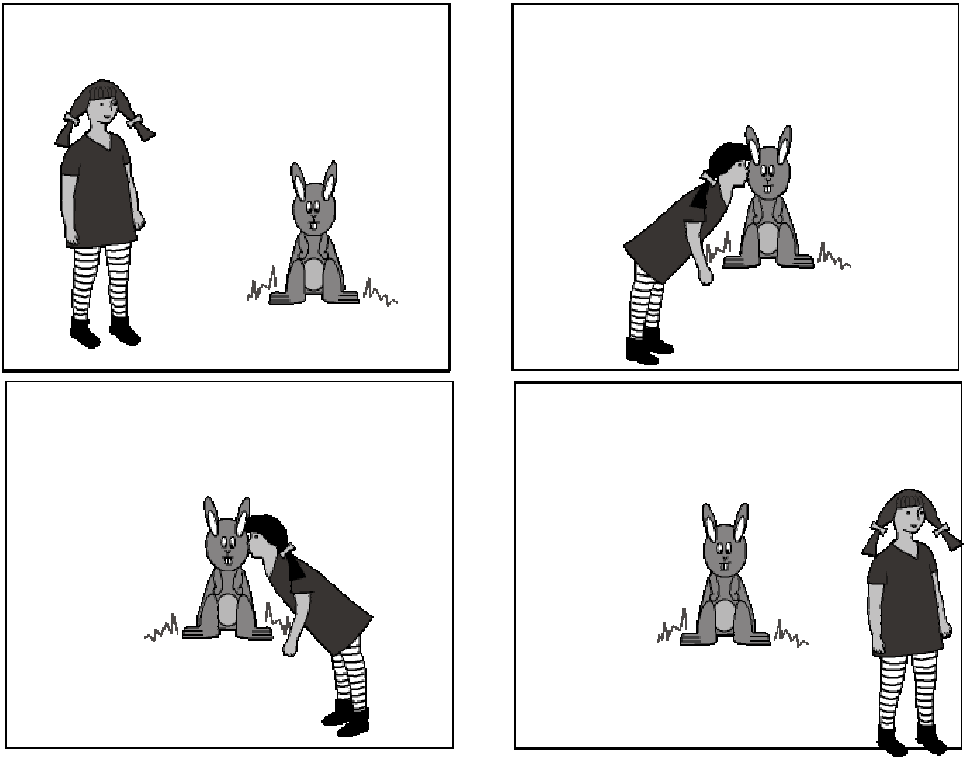
Picture 1 Dit is een meisje, dit is een hond en dit zijn paarden.
Een zwarte, een grijze en een witte. Die wil ze denk ik aaien.
'This is a girl, this is a dog and these are horses.
A black one, a grey one and a white one. I think shes wants to stroke them.'

Picture 2 Hier aait ze de hond.
'Here, she's stroking the dog

Picture 3 En hier aait ze een paard.
'And here she's stroking a horse.'

Picture 4 En nu gaat ze weer weg.
'And now she's leaving.'

Puppet:	a.	Het meisje	heeft	een paard	twee keer	geaaid	[target: rejection]
		the girl	has	a horse	two times	stroked	
		'The girl stroked a (certain) horse twice.'					
	b.	Het meisje	heeft	twee keer	een paard	geaaid	[target: rejection]
		the girl	has	two times	a horse	stroked	
		The girl stroked a(ny) horse twice.'					



(16) Example scenario for 'yes-reponse' filler item in truth value judgement task on scrambling across *twee keer*

Picture 1 Dit is een meisje, dit is een konijn. Ze wil het denk ik zoenen.
'This is a girl and this is a rabbit. I think shes wants to kiss it.'

Picture 2 Hier zoent ze het konijn.
'Here, she's kissing the rabbit.'

Picture 3 En hier zoent ze het konijn.
'And here she's kissing the rabbit.'

Picture 4 En nu gaat ze weer weg.
'And now she's leaving.'

Puppet:

a.	Het meisje	heeft	het konijn	twee keer	gezoend	[target: acceptance]
	the girl	has	the rabbit	two times	kissed	

'The girl kissed a (certain) rabbit twice.'

b.	Het meisje	heeft	twee keer	het konijn	gezoend	[target: acceptance]
	the girl	has	two times	the rabbit	kissed	

The girl kissed a(ny) rabbit twice.'

All subjects heard the same pseudo-random order. A blocked design was used, that is, the scrambled test items interspersed with the scrambled fillers were presented first, followed by the non-scrambled test items with the non-scrambled fillers. There were several reasons for using such a blocked design rather than mixing up the scrambled and non-scrambled orders with each other. First of all, alternating scrambled and non-scrambled orders might draw subjects' – especially the older L2ers' – attention to the difference between the two forms and this could potentially lead them to realise that this was the focus of the task. By presenting the different test items in blocks, with fillers which had the same order, this possible effect can be controlled for. Secondly, the difference between the scrambled and non-scrambled form is rather subtle; adult native speakers often do not even hear the difference. An informal pilot study suggested that when the two orders are mixed and, in particular, when a non-scrambled form is followed by a scrambled form, adult native speakers find this slightly odd from a discourse point of view. Although they will often (but not always) still provide targetlike responses, such a presentation order induces several raised eyebrows and/or confused looks. In the design adopted here, then, if there are carry-over effects from scrambled-item response to non-scrambled-item response, this should be detectable at the 'changeover point' from scrambled to non-scrambled items in the middle of the task. The scrambled order was presented before the non-scrambled order because it is the most important condition. By presenting this first, we can be certain that if subjects provide targetlike responses, it is not because of some task-specific strategy. Imagine, by contrast, that (some) subjects were to hear the non-scrambled order first. If they accepted the non-scrambled order (as would be expected, given that it is consistent with L1 transfer and it is targetlike), at the point at which they heard the scrambled order, they might reason that because the scrambled form sounds 'different' or 'odd', or because they do not know what it means, they should answer differently and, consequently, they might reject it. This would result in a response on the test items which is targetlike, but for non-targetlike reasons. In the chosen design, the scrambled and non-scrambled items directly adjacent to each other at the changeover point are fillers. The pictures for both test items and fillers were in principle the same in both conditions, but the clothing, colours, etc. were changed in order to add variety and to distract.

As in Krämer's original experiment, the intonation used for all items was the most natural for the scrambled order, namely with the pitch rise for the VP starting on *twee keer* 'twice'. The indefinite determiner was always pronounced as *een* \ən\ rather than *één* 'one' (see §2.1.2 and §5.4.3.3 for relevant discussion). This was also the case for the non-scrambled items.

6.4.1.2 Procedure

The truth value judgement task with *twee keer* was carried out using exactly the same procedure as for the truth value judgement task with negation. The *twee* *keer* task was the first to be carried out. It was preceded by a brief training session, as detailed in §6.3.1.2, and this in turn was preceded by a pre-test, taken from Krämer (2000), to ensure that subjects understood the meaning of *twee keer*. Although it is highly likely that the youngest subjects (7-year-olds – although, in principle, they could have been as young as four or 5) know how to count, the pre-test was still included to (a) double-check that this was the case and (b) check that subjects knew the target-language lexical item. In the pre-test, subjects were presented with a set of farm animals, namely a pig, a horse and a sheep, and a small fence. The experimenter manipulated the objects in the following way: in the first item, the pig jumped over the fence twice and the experimenter asked how often the pig jumped over the fence. In the second item, the horse jumped over the fence three times and the pig twice and the subject was asked to say who jumped over the fence three times. In the final item, the sheep jumped twice and the pig three times, and the subject was asked how many times the sheep jumped over the fence. Subjects had to pass this test in order to proceed to the experimental task. Only one (child L2) subject failed at this stage.

6.4.2 L1 acquisition

The purpose of carrying out the truth value judgement task with L1 children was threefold. First, given that the 4- to 7-year-old children in the act-out task with *twee keer* in Krämer's study provided on average 48.8% (66/135) targetlike responses to sentences containing indefinite objects scrambled across *twee keer*, it is necessary to test older children in order to establish the age at which L1 children reach adultlike levels. Second, using the same type of task for scrambling across negation and for scrambling across *twee keer* will eliminate the possibility of any task effect, that is, it will be possible to make a more straightforward comparison of the development of scrambling in these two cases. Third, as noted above with respect to the negation experiment, in order to have L1 data which are maximally comparable to the L2 data, it is essential to carry out with the L1 children exactly the same experiment used with the L2ers. It should be noted that the within-subjects design of the present study contrasts with the between-subjects design employed by Krämer; a within-subjects design allows us to determine whether individual children differentiate between scrambled and non-scrambled indefinites. The individual results are given in §6.4.2.2. First, the group results are presented.

6.4.2.1 Group results

One child (L112#5C) was excluded because he failed on more than one filler. Of the remaining fillers, 99.4% (1559/1568) were correct. The results for the test items for each of the six age groups, plus adult controls, are presented in Table 46.

Table 46. L1 children (comprehension):
Group results on *twee keer* task. Average percentage of targetlike responses

Subjects		No scrambling condition (target: no)		Scrambling condition (target: yes)	
Age group	**n**	**%**	**SD**	**%**	**SD**
7-year-olds	14	81.4	38.0	25.7	41.8
8-year-olds	16	96.0	15.5	18.7	38.9
9-year-olds	15	86.7	29.0	56.0	40.1
11-year-olds	15	88.0	27.0	60.0	46.0
12-year-olds	14	94.3	17.1	94.3	22.2
13-year-olds	25	---	---	98.0	10.0
Adults	14	95.7	11.6	90.0	28.0

Let us first start with the no scrambling condition. More or less all of the child L1 groups consistently accept the non-scrambled utterance in the context where two objects are manipulated, as do the adults. There is no significant difference between groups in this condition (ANOVA: df = 5, F = 1.224, p = .305). In the scrambling condition, the youngest two groups of children, the 7- and 8-year-olds, regularly give non-targetlike responses, that is, they accept sentences containing an indefinite NP object scrambled across *twee keer* in a context which is incompatible with a specific reading for that NP. The oldest two groups, the 12- and 13-year-olds, consistently reject the non-specific reading for the scrambled indefinite object, however, and the 9- and 11-year-olds' responses are targetlike in just more than half of the items. The motivation the children provide for their targetlike answers, exemplified in (17), indicates that they attribute a specific interpretation to the scrambled indefinite object NP.

(17)	Test sentence:	Het meisje	heeft	een bal	twee keer	gegooid	
		the girl	has	a ball	twice	thrown	
		'The girl threw a (certain) ball twice.'					
	Subject L111#9C:	Nee,	ze	heeft	twee verschillende ballen		gegooid
		no	she	has	two different balls		thrown
		'No, she threw two different balls.'					

In this condition, there is a statistically significant difference between groups (ANOVA: df = 6, F = 16.483, p = .000). Post-hoc Games-Howell analyses were carried out in order to determine which groups differed from each other. The results are given in Table 47.

Table 47. L1 children (comprehension): Results of post-hoc Games-Howell test on *twee keer* task (scrambling condition)

Comparison		**Mean difference %**	**Significance**
7-**year-olds**	8-year-olds	-8.2	p = .993
	9-year-olds	-30.3	p = .376
	11-year-olds	-34.3	p = .316
	12-year-olds	-68.6	p = .000
	13-year-olds	-72.3	p = .000
	adults	-64.3	p = .001
8-year-olds	9-year-olds	-38.5	p = .097
	11-year-olds	-42.5	p = .088
	12-year-olds	-76.8	p = .000
	13-year-olds	-80.5	p = .000
	adults	-72.5	p = .000
9-year-olds	11-year-olds	-4.0	p = 1.000
	12-year-olds	-38.3	p = .039
	13-year-olds	-42.0	p = .013
	adults	-34.0	p = .151
11-year-olds	12-year-olds	-34.3	p = .142
	13-year-olds	-38.0	p = .060
	adults	-30.0	p = .365
12-year-olds	13-year-olds	-3.7	p = .988
	adults	4.3	p = .999
13-year-olds	adults	8.0	p = .938

Both the 7- and the 8-year-olds differ significantly from the 12- and 13-year-olds and the adults. The 9-year-olds also differ significantly from the 12- and the 13-year-olds, but not from the adults or the 11-year-olds. The difference between the 8- and the 9-year-olds and between the 8- and the 11-year-olds is approaching significance, as is the difference between the 11- and the 13-year-olds. The 11-, 12- and 13-year-olds do not differ from the natives and the 12- and 13-year-olds do not differ from each other.

6.4.2.2 Individual results

The individual results from each subject are provided in Appendix D. As with the negation task, subjects were categorised according to whether they showed acceptance, rejection or mixed patterns (see §6.3.2.2 for details). The distribution of L1 children and adults across these three patterns in the scrambling condition is given in Figure 6.

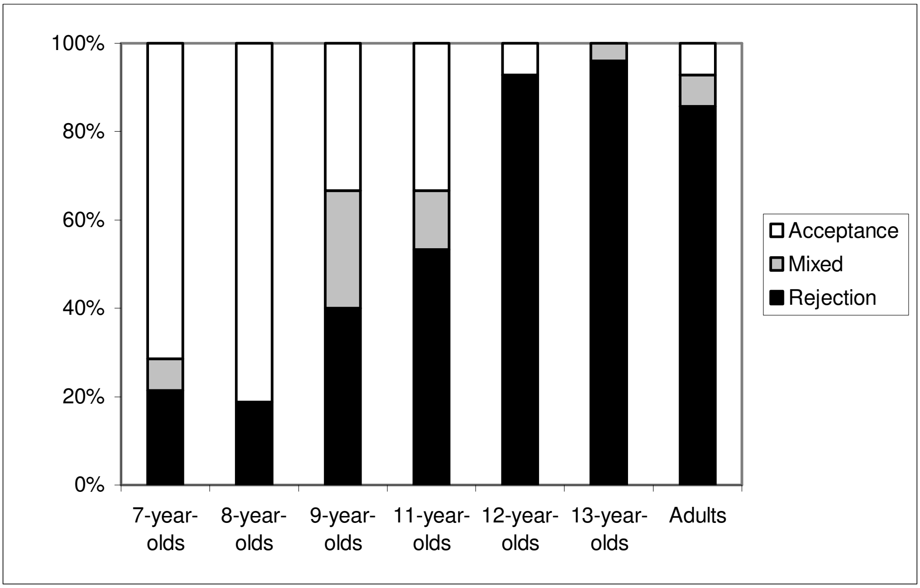


Figure 6. L1 children (comprehension) : Distribution of individual response patterns on *twee keer* task (scrambling condition)

Let us begin with the adult controls. Most of the adults (12/14) reject the non-specific interpretation for the indefinite object NP scrambled across *twee keer* as expected. Contrary to expectation, however, one adult consistently accepts the scrambled test sentences. Initially, this subject spontaneously rejected the first test item, as expected, but subsequently, during the following filler item, he decided that if the NP was not realised as *één*, the scrambled sentence was acceptable after all. The production data from the native adults presented in Chapter 5 indicate that *één* is indeed the preferred realisation of the scrambled indefinite object (cf. §5.4.3.3). Nevertheless, the observation that the vast majority of adult native speakers reject the non-specific reading for the scrambled indefinite when it is *not* realised as *één* demonstrates that it is the position rather than the form of the NP which is associated with the specific meaning. It seems that this particular subject was applying metalinguistic knowledge in this experiment.[20] The remaining adult subject rejects three scrambled test items and accepts two. This could be considered as noise.

Turning to the L1 children, the majority of children in the youngest two groups have a (non-targetlike) pattern of acceptance. At the other end of the spectrum, virtually all the children in the oldest two groups have a (targetlike) pattern of rejection. In the 9- and 11-year-old groups, the majority of subjects have a pattern of rejection or a mixed pattern, but still a not insignificant number of subjects has a pattern of acceptance. Broadly speaking,

[20] This subject consistently made targetlike judgements in the negation task, even though there, too, the scrambled indefinite object was not realised as *één*.

there appear to be three different groups of children: the 7- and 8-year-olds, who are generally non-targetlike, the 9- and 11-year-olds who are more targetlike than the youngest children but still not at adultlike levels and the 12- and 13-year-olds who are more or less consistently targetlike.

Let us now consider the no scrambling condition (exact numbers available in Appendix D). In all groups, the vast majority of subjects have a targetlike pattern of acceptance. There are handful of subjects (12.2% (9/74) across all groups) who occasionally or consistently reject the non-specific meaning for the non-scrambled order.[21] They form a subset of the subjects who reject the non-specific meaning of the scrambled form. Recall that all the scrambled items were presented *before* the non-scrambled items, in a block with scrambled fillers. It appears that when in the no scrambling condition these subjects were presented with a story similar to one they had previously heard in the scrambling condition, they sometimes responded in the same way. In other words, it is only those subjects who reject the scrambled items who also reject the non-scrambled items, that is, that there are no subjects who reject the non-scrambled items and accept the scrambled items. This observation suggests that we are dealing with a carry-over effect. Comparing responses to the two conditions for the other subjects, we find that 31 out of the remaining 65 children (47.7%) make a relative distinction between the two, that is, they consistently accept the non-specific interpretation for the non-scrambled form and they reject this reading for the scrambled form some or all of the time. There are 24 children who make an absolute distinction between the two conditions. As expected, the number of children who differentiate between the two conditions increases with age. Those making a relative distinction include one 7-year-old, two 8-year-olds, nine 9-year-olds, nine 11-year-olds and 12 12-year-olds and the number of children making an absolute distinction, respectively, are zero, two, four, six and 11, respectively.

6.4.2.3 Discussion

The results presented here for the scrambled indefinite object NP show that 7- and 8-year-old L1 Dutch children generally make non-targetlike judgements, accepting the non-specific interpretation depicted in the experimental scenario. There is a leap towards targetlike behaviour at age 9, and this seems to plateau, until age 12, when children are no longer distinguishable from adults.

The gradual pattern of development over several years observed in the data here are consistent with Philip's (2003) data. These two sets of data differ, however, in terms of the proportion of targetlike subjects amongst the older children. In Philip's oldest group, the 12-year-olds, only 44.2% (19/43) of children provided targetlike responses. This contrasts with the 92.9% (13/14) targetlike 12-year-olds tested here. There are (at least) two differences between

---

[21] The 13-year-olds are not included in this discussion as they were tested on the scrambling condition only.

the task used here and the one used by Philip which might explain this discrepancy. First of all, in Philip's study, each child was tested on one item only per condition, whereas the children here heard five items in each condition. Philip's reason for this methodological choice is the possibility that a repeated measures design might lead to a practice effect, that is, subjects' responses to a second item will be affected by their response to the first item, and their responses to a third item by their responses to the first and second items, and so on. As noted in §3.3.1.2, it is not entirely clear what such a practice effect should look like, but for the sake of argument, let us suppose that this difference in methodology might explain the difference between these two sets of results. If this were the case, we might expect that once the first item only is taken into account, the number of targetlike subjects will be considerably lower. This is not the case, however. Although using only the first item means that the proportion of targetlike subjects is slightly depressed for the 9-, 11- and 13-year-olds, the overall pattern remains the same, and in general, the percentages of targetlike subjects based exclusively on the first item are only marginally different from those using all five items. Compare the following percentages with those in Figure 6: 28.6% (4/14) of 7-year-olds, 18.8% (3/16) of 8-year-olds, 40% (6/15) of 9-year-olds, 46.7% (7/15) of 11-year-olds, 92.9% (13/14) of 12-year-olds and 84% (21/25) of 13-year-olds. The difference in number of items per subject can therefore not explain the discrepancy between the two sets of results. A second difference between Philip's task and the present task is that in Philip's task, the test sentences were embedded in a more complex story. In another experiment, where Philip compares children's responses on this complex story with a much simplified version of the same story, he observes that children who hear the simpler story are more targetlike than those who hear the complex story: compare 40% of the 43 children (aged 4;4 – 7;0) correctly rejecting the non-specific interpretation of the scrambled indefinite in the simpler story to just 8% of the 51 children (aged 3;5 – 7;9) doing so in the more complex story. As Philip notes, this result is consistent with Krämer's Non-Integration Hypothesis, that is, when there is more discourse for children to integrate, as in the more complex story, fewer children make targetlike judgements. The stories employed in the present task were much simpler than those in Philip's task, and thus, in a comparable fashion to Philip's simple-complex story comparison, this could plausibly account for the higher proportion of targetlike subjects.

The proportion of targetlike subjects on the current task is slightly lower than on Krämer's act-out task. There, 39.1% (9/23) of the 7-year-olds consistently had a specific reading for the scrambled indefinite object, while here, three of the 14 subjects (21.4%) in this age group do. As with the child L1 negation data, the child L1 *twee keer* data, which all show that development is gradual and relatively late, are compatible with Krämer's results and her Non-Integration Hypothesis.

6.4.3 L2 acquisition

The purpose of carrying out this task with L2 children and adults is the same as for the previous task, namely to investigate the possible effect of limited discourse integration in child L2 acquisition and its consequences for the child L2 ~ adult L2 comparison, as well as to determine whether there are any English-speaking L2ers who have targetlike knowledge of the interpretive constraints on scrambling. Exactly the same task was used with the L2 subjects as with the L1 children. The group results are presented first in §6.4.3.1, followed by the individual results in §6.4.3.2.

6.4.3.1 Group results

Virtually all of the fillers were answered correctly (99.0% (578/584)), and hence none of the subjects had to be excluded. As above, the results are first presented for each proficiency level and subsequently an analysis according to age at time of testing is also carried out.

6.4.3.1.1 Proficiency level

As with the negation experiment, the analysis in this section serves to determine whether the various proficiency levels differ from each other (within and across the two L2 groups), whether they differ from the native adults, and whether there is any evidence for development in targetlike knowledge of scrambling with increasing proficiency. The no scrambling condition will be considered first. The results are given in Table 48.

Table 48. L2 children and adults (comprehension): Group results (organised by proficiency) on *twee keer* task. Average percentage of targetlike (i.e. yes) responses in no scrambling condition

	L2 children			L2 adults			Mann-Whitney tests comparing child/adult groups
	n	**%**	**SD**	**n**	**%**	**SD**	
Low	14	91.4	26.8	18	97.8	6.5	Z = -.331, p = .741
Mid	6	90.0	16.7	8	80.0	35.5	Z = .302, p = .763
High	16	86.7	29.5	11	90.9	30.2	Z = -1.474, p = .140

All of the groups consistently accept the non-specific reading for the non-scrambled test sentence, although the mid proficiency L2 adults do so at a slightly lower rate than the other groups. Note that this group also contains the most variation. The reason for the (unexpected) non-targetlike responses in this condition will be discussed below in §6.4.3.2, where the individual data are presented. The results for the Mann-Whitney tests given in the final column of Table 48 indicate that in this condition, the L2 children and adults do not differ from each other in any of the three proficiency levels. Comparing the different proficiency levels with each other and with native adults, no significant differences were observed for either the L2 children (ANOVA: df = 2, F = .573, p = .569) or the L2 adults (Kruskal-Wallis: χ^2 = 3.725, df = 3, p = .293).[22]

We now turn to the scrambling condition. Recall that in this condition, a targetlike response involves rejecting the non-specific interpretation depicted in the story for the indefinite object NP scrambled across *twee keer*. Table 49 presents the average percentage of targetlike responses per proficiency group for L2 children and L2 adults.

Table 49. L2 children and adults (comprehension): Group results (organised by proficiency) on *twee keer* task. Average percentage of targetlike (i.e. no) responses in scrambling condition

	L2 children			L2 adults			Mann-Whitney tests comparing child/adult groups
	n	%	SD	n	%	SD	
Low	14	7.1	26.7	18	31.1	37.7	Z = -2.583, p = .010
Mid	6	53.3	45.0	8	25.0	41.1	Z = -1.110, p = .267
High	16	50.0	46.8	11	63.6	43.7	Z = -.756, p = .449

Taking all groups together, the average rate of targetlike responses ranges from 7.1% to 63.6%. The L2ers motivate their targetlike responses in the same way as the L1 children, referring to the fact that two different objects were manipulated rather than one and the same object:

(18) a. Test sentence: Het meisje heeft een teddybeer twee keer gezoend
the girl has a teddy bear two times kissed
'The girl kissed a (certain) teddy bear twice.'

Subject A26C: Nee ze heeft twee verschillende teddys gezoend
no she has two different teddys kissed
'No, she kissed two different teddy bears.'

b. Test sentence: De jongen heeft een hond twee keer geaaid
the boy has a dog two times stroked
'The boy kissed a (certain) dog twice.'

[22] The Kruskal-Wallis test is a non-parametric test using ordinal data which can be used when an ANOVA is not possible. The L2 data were thus converted to an ordinal scale for this purpose.

Subject C24C:

Nee,	omdat	hij	heeft	de bruine	één keer	geaaid
no	because	he	has	the brown	one time	stroked

'No, because he stroked the brown one once and …

en	de witte	één keer	geaaid,
and	the white	one time	stroked

.. the white one once ….

niet	de bruine	twee keer	geaaid
not	the brown	two times	stroked

.. not the brown one twice.'

It is unlikely that subjects' non-targetlike responses resulted from their failure to notice that two different objects were manipulated, as this aspect of the stories was often commented upon:

(19) a. Test sentence:

Het meisje	heeft	een bal	twee keer	gegooid
the girl	has	a ball	twice	thrown

'The girl threw a ball twice.'

Subject A17C:

Ja,	een bal	maar	er	zijn	twee	verschillende	ballen
yes	a ball	but	there	are	two	different	balls

'Yes, a ball, but there are two different balls.'

Maar	ja,	ik	zou	zeggen	dat	dat	klopt
but	yes	I	would	say	that	that	is-right

'But yes, I would say that that's right.'

b. Test sentence:

Het meisje	heeft	een aap	twee keer	gekieteld
the girl	has	a monkey	two times	tickled

'The girl tickled a (certain) monkey twice.'

Subject A15C:

Ja,	eerst	de bruine	en	dan	de rode
yes	first	the brown	and	then	the red

'Yes, first the brown one and then the red one.'

As on the negation task (but not as often), subjects' responses occasionally appeared to be affected by the form of the NP (*een* vs. *één*), as exemplified in (20). As noted in the discussion of the negation task, this could be interpreted as subjects associating the specific interpretation with the cardinal/stressed indefinite *één* rather than the position of the NP. On the other hand, however, given that subjects never made such comments in response to the non-scrambled test items, these responses could be taken to indicate that subjects are aware of the preferred realisation of the scrambled indefinite as *één* rather than *een*.

(20) Test sentence:

Het meisje	heeft	een bal	twee keer	gegooid
the girl	has	a ball	twice	thrown

'The girl threw a ball twice.'

Subject A23C:	Ze	heeft	twee ballen	gegooid			
	she	has	two balls	thrown			
	'She threw two balls.'						
	Dus	het	is	niet	één bal	maar	twee ballen
	so	it	is	not	one ball	but	two balls
	'So it's not one ball but two balls…						
	maar	ze	heeft	twee keer	gegooid,	ja	
	but	she	has	two times	thrown	yes	
	…but she threw two times, yes.'						

Given that the subject's response is unclear, the experimenter asks whether she should repeat the sentence, the subject says yes and subsequently responds:

Een \ən\ bal,	ja.
A ball,	yes
'A ball, Yes.'	

As pointed out to me by Irene Krämer (p.c., 16 November, 2004), however, subject A23C's comments in (20) could also be interpreted as meaning 'the sentence isn't right as far as the number of balls is concerned, but it is ok with respect to the number of times which a throwing action took place.' On such an interpretation, this response should be interpreted as targetlike. In the results presented below, it is the subject's final answer which is counted, that is, the response in (20) is counted as non-targetlike. This is also the case for subject C25C whose initial response in one item could potentially be seen as a targetlike rejection of the test sentence but who nevertheless goes on to accept it ((21)-a) and, for another subject, A17C, who explicitly states that she is not sure about her response ((21)-b).[23]

(21)	a.	Test sentence:	Het meisje	heeft	een aap	twee keer	gekieteld						
			the girl	has	a monkey	two times	tickled						
			The girl tickled a (certain) monkey twice.'										
		Subject C25C:	Ze	heeft	TWEE apen	gekieteld							
			she	has	two monkeys	tickled							
			'She tickled TWO monkeys.'										
		Puppet:	Dus	heb	ik	het	goed	of	heb	ik	het	fout?	
			so	have	I	it	right	or	have	I	it	wrong	
			'So am I right or am I wrong?'										

[23] Even though she is aware of the ambiguity in her English, subject A17C clearly does not know the interpretive restrictions on scrambled indefinites in Dutch. Subjects A23C and C25C might know them, in which case the child L2 and adult L2 high proficiency groups would have one more targetlike subject each.

Subject C25C:	Dus	heb	je	het	volgens	mij	goed
	so	have	you	it	following	me	right
	'(hesitating) So I think you've got it right.'						

b.	Test sentence:	De jongen	heeft	een varken	twee keer	geknepen
		the boy	has	a pig	two times	pinched
		'The boy pinched a (certain) pig twice.'				

Subject A17C: I've no idea whether it's ambiguous in Dutch as well, but anyway, I'd say yes.

The Mann-Whitney tests presented in the final column in Table 49 indicate that there is no significant difference between the mid proficiency L2 children and adults, nor between the high proficiency L2 children and adults, but there is a difference between the two low proficiency groups.

Once again, the three proficiency groups were compared with each other and the native adults. A significant difference was observed for both the L2 children (ANOVA: df = 3, F = 11.834, p = .000) and the L2 adults (ANOVA: df = 3, F = 8.478, p = .000). The results of the post-hoc analyses, carried out to determine which proficiency groups differed from each other and/or from the native adults for the L2 children and for the L2 adults, respectively, are presented in Table 50, where significant differences are highlighted.

Table 50. L2 children and adults (comprehension): Results of Post-hoc Games Howell test on *twee keer* task (scrambling condition)

Comparison	Mean difference %	Significance
Child L2 Low vs. Child L2 Mid	-45.9	p = .182
Child L2 Low vs. Child L2 High	-43.2	p = .022
Child L2 Mid vs. Child L2 High	3.2	p = .999
Child L2 Low vs. Adult natives	-83.4	p = .000
Child L2 Mid vs. Adult natives	-37.4	p = .331
Child L2 High vs. Adult natives	-40.0	p = .038
Adult L2 Low vs. Adult L2 Mid	6.8	p = .983
Adult L2 Low vs. Adult L2 High	-33.3	p = .206
Adult L2 Mid vs. Adult L2 High	-39.1	p = .240
Adult L2 Low vs. Adult natives	-59.4	p = .000
Adult L2 Mid vs. Adult natives	-64.7	p = .010
Adult L2 High vs. Adult natives	26.1	p = .336

Within the child L2 group, the low proficiency group differs significantly from the high proficiency group and the native adults, and the high proficiency group also differs from the native adults. As for the adult L2 group, there are no significant differences between the three proficiency groups, although the mean differences indicate that the rate of targetlike responses in the low and mid groups is different from that of the high proficiency group. Both the low and mid proficiency adult L2 groups differ significantly from the natives, but the high proficiency adult L2 group does not.

When the responses on the scrambling and no scrambling conditions are compared with each other, it is only the low proficiency child and adult groups who make a significant distinction between the two (Child Low: t = -5.916, p = .000; Adult Low: t = -6.689, p = .000). The difference for the mid proficiency level adults is also approaching significance (Adult Mid: t = - .128, p = .071). The other groups do not make a significant difference between the two conditions (Child Mid: t = -1.611, p = .168; Child High: t = - 1.448, p = .168; Adult High: -1.520, p = .160). In other words, there is an effect of proficiency: the least proficient child/adult L2 subjects have on average significantly more targetlike responses in the no scrambling condition than in the scrambling condition, but this is no longer the case for the higher proficiency child/adult L2 subjects.

For the L2 children, a significant moderate correlation was found between targetlike responses in the scrambling condition and length of exposure (r = .530, p = .001) and age at time of testing (r = .517, p = .001) and a weak-moderate correlation between targetlike responses and proficiency score (r = .456, p = .007). For L2 adults, there were significant weak correlations between targetlike responses and length of exposure (r = .420, p = .01) and age at time of testing (r = .362, p = .028), but no correlation with proficiency score (r = .256, p = .165). A regression analysis with proficiency and age at time of testing (for L2 children) shows that age is the best predictor (r^2 = .244, p = .003). However, when length of exposure is included in the analysis, this is the best predictor for targetlike responses in the *twee keer* scrambling condition (r^2 = .257, p = .002 for the L2 children and r^2 = .163, p = .024 for the L2 adults). Note, however, that the amount of variance which these predictor variables explain in the data is minimal at 25.7% for the L2 children and 16.3% for the L2 adults for length of exposure and 24.4% for age at time of testing for the L2 children (cf. almost 50% of variation is explained by proficiency in the L2 production data).

6.4.3.1.2 Age at time of testing

In order to determine whether an age effect similar to the one observed in child L1 acquisition is found in L2 acquisition, the child L2 data are now examined in terms of the children's age at time of testing. The results for both the scrambling and no scrambling conditions are given in Table 51.

Table 51. L2 children (comprehension): Comparison of two conditions in *twee keer* task (group results organised by age at time of testing). Average percentage of targetlike responses

Subjects		**No scrambling condition (target: yes)**		**Scrambling condition (target: no)**	
Age group	**n**	**mean %**	**SD**	**mean %**	**SD**
7-year-olds	3	100.0	0	0	0
8-year-olds	10	98.0	6.3	0	0
9-year-olds	3	60.0	52.9	86.7	23.1
10-year-olds	5	75.0	50.0	52.0	50.2
11-year-olds	5	88.0	26.8	32.0	46.0
13/14-year-olds	8	90.0	15.1	67.5	42.7

On the whole, each of the six child L2 age groups correctly accept the non-specific interpretation for the non-scrambled form. There is no significant difference between the different age groups and the native adults (Kruskal-Wallis: df = 6, χ^2 = 7.364, p = .289). The 9- and 10-year-olds do have a considerably lower rate of acceptance than the other groups, however. These two groups also contain considerably more variation than the other groups; it is possible that there are one or two subjects who are responsible for the depressed acceptance rate. This will be addressed in the next section.

Turning to the scrambling condition, the four oldest child L2 groups provide between on average 32.0% and 86.7% targetlike responses. This contrasts sharply with the youngest two groups, the 7- and 8-year-olds, who produce no targetlike responses whatsoever, that is, they consistently accept the non-specific reading for the indefinite object NP scrambled across *twee keer*. Comparing the various age groups with each other and the native adults, a significant difference is observed (ANOVA: df = 6, F = 11.449, p = .000). The results of the post-hoc analyses, showing which groups differ from each other, are given in Table 52.

Table 52. L2 children (comprehension): Results of Post-hoc Games Howell tests for *twee keer* task (scrambling condition)

Comparison		Mean difference %	Significance
7-year-olds	8-year-olds	0	
	9-year-olds	-86.7	p = .080
	10-year-olds	-52.0	p = .355
	11-year-olds	-32.0	p = .659
	13/14-year-olds	-67.5	p = .022
	native adults	-90.0	p = .000
8-year-olds	9-year-olds	-86.7	p = .080
	10-year-olds	-52.0	p = .355
	11-year-olds	-32.0	p = .659
	13/14-year-olds	-67.5	p = .022
	native adults	-90.0	p = .000
9-year-olds	10-year-olds	34.7	p = .764
	11-year-olds	54.7	p = .344
	13/14-year-olds	19.2	p = .920
	native adults	-3.3	p = 1.000
10-year-olds	11-year-olds	20.0	p = .982
	13/14-year-olds	-15.5	p = .990
	native adults	-38.0	p = .689
11-year-olds	13/14-year-olds	-35.5	p = .732
	native adults	-58.0	p = .274
13/14-year-olds	native adults	-22.5	p = .822

The 7- and the 8-year-olds differ significantly from the 13/14-year-olds and the native adults. The difference between these younger groups and the 9-year-olds is also approaching significance. No other significant differences are observed, that is, there are no differences between the 9-, 10-, 11- and 13/14-year-old groups, nor between these and the native adults.

As mentioned in §6.3.3.1.2, a comparison of the proficiency levels of the various child L2 age groups is complicated by the different number of subjects and the different range of proficiency scores within each group. Also, because the number of targetlike responses is correlated with proficiency and the oldest L2 children are generally more proficient, there is little variation in terms of scrambling responses in this oldest group. This is unfortunate because this is the only group which contains children who are as old as or older than the child L1 groups which are consistently targetlike, that is, the 12- and 13-year-olds. In other words, to tease apart the factors of age and transfer, it would be necessary to have a larger group of older L2 children at different proficiency levels.[24] A close examination of the data in this group as well as the

[24] The problem with the current data is that all of the subjects in the oldest group (the 13/14-year-olds) were first exposed to Dutch at either age four or age five. There is no *a priori* reason why this should be the case. Following the criterion adopted for child L2 acquisition, it

youngest two groups reveals some interesting patterns, however. In the 13/14-year-old group, there are two non-targetlike subjects (out of eight); one (C35C) has the lowest proficiency score in the group and the other (C10C) the fourth lowest. Amongst the 15 children in the youngest two groups, four have a proficiency score which is higher than the lowest proficiency score for the targetlike older children and, nevertheless, these four provide non-targetlike responses.

As with the negation task, I now consider the adult L2 group in terms of age at time of testing. The purpose of this analysis is to factor out age at time of testing for this group. This is important because it is possible that the adult L2 group's younger subjects who are in the same age range as non-targetlike L1 children are affecting the adult L2 group's score as a whole. Once again, the adult L2 group will be divided into a younger group (age at time of testing: 9;8 – 14;11) and an older group (age at time of testing: 22;0 – 50;0). The results for the no scrambling condition are given in Table 53.

Table 53. Younger and older L2 adults (comprehension): Group results (organised by proficiency) on *twee keer* task. Average percentage of targetlike (i.e. yes) responses in no scrambling condition

	L2 adults (younger subjects only)			L2 adults (older subjects only)			L2 adults (original)		
	n	**%**	**SD**	**n**	**%**	**SD**	**n**	**%**	**SD**
Low	12	98.3	5.8	6	96.7	8.2	18	97.8	6.5
Mid	3	66.7	57.7	5	88.0	17.9	8	80.0	35.5
High	3	100.0	0	8	87.5	35.4	11	90.9	30.2

The only observable differences between the younger and older adult L2 groups is that at mid proficiency level, the younger adults have a lower rate of acceptance than the older adults, and at the high proficiency level, the younger L2 adults are more targetlike than the older L2 adults. There is no significant difference between the three proficiency levels and the native adults for the younger L2 adult group (Kruskal-Wallis: df = 3, χ^2 = 2.193, p = .533), nor for the older L2 adult group (Kruskal-Wallis: df = 3, χ^2 = 1.712, p = .634).

The results in the scrambling condition for these two groups are presented in Table 54.

would in principle be possible to have subjects in this group who were first exposed to Dutch at age six or seven. Given that length of exposure generally correlates with proficiency, it would be expected that such subjects would be less proficient than those with an earlier age at first exposure, which would lead to more variation within the oldest child L2 group. This, in turn, would enable a potentially more fruitful investigation of the interaction between these different variables. As noted in §6.2, my endeavours to locate such subjects were unfortunately without success.

Table 54. Younger and older L2 adults (comprehension): Group results (organised by proficiency) on *twee keer* task. Average percentage of targetlike (i.e. no) responses in scrambling condition

	L2 adults (younger subjects only)			L2 adults (older subjects only)			L2 adults (original)		
	n	%	SD	n	%	SD	n	%	SD
Low	12	23.3	31.7	6	46.7	46.8	18	31.1	37.7
Mid	3	26.7	46.2	5	24.0	43.3	8	25.0	41.1
High	3	26.7	46.2	8	77.5	36.1	11	63.6	43.7

In this condition, teasing apart the younger and older L2 adults into two separate groups has a clear effect: in both the low and high proficiency older adult L2 groups, the percentage of targetlike responses is considerably greater than in the equivalent younger adult L2 groups. In other words, it is the older adult L2 subjects who are responsible for most of the targetlike responses in the original adult L2 group. There is no significant difference between the three proficiency levels in either the older (Kruskal-Wallis: $\chi^2 = 4.002$, df = 2, p = .135) or the younger (Kruskal-Wallis: $\chi^2 = .042$, df = 2, p = .979) adult L2 groups. When the native adults are included in this comparison, however, there is a significant difference for both the older (ANOVA: F = 5.081, df = 3, p = .006) and the younger (ANOVA: F = 7.647, df = 3, p = .001) adult L2 groups. The results of the post-hoc tests in Table 55 show that the older adult L2 mid proficiency group significantly differs from the native controls and almost from the older adult L2 high proficiency group. Within the younger adult L2 group, the low proficiency group significantly differs from the native controls, and the differences between the mid proficiency group and the natives and between the high proficiency group and the natives are almost significant.

Table 55. Younger and older L2 adults (comprehension): Results of Bonferroni post-hoc test on *twee keer* task (scrambling condition)

Comparison	**Mean difference %**	**Significance**
Younger Adult Low vs. Younger Adult Mid	-3.3	p = 1.000
Younger Adult Low vs. Younger Adult High	-3.3	p = 1.000
Younger Adult Mid vs. Younger Adult High	0	p = 1.000
Younger Adult Low vs. Natives	-63.9	p = .001
Younger Adult Mid vs. Natives	-60.6	p = .071
Younger Adult High vs. Natives	-60.6	p = .071
Older Adult Low vs. Older Adult Mid	22.7	p = 1.000
Older Adult Low vs. Older Adult High	-30.8	p = .748
Older Adult Mid vs. Older Adult High	-53.5	p = .087
Older Adult Low vs. Natives	-43.3	p = .121
Older Adult Mid vs. Natives	-66.0	p = .009
Older Adult High vs. Natives	-12.5	p = 1.000

Interestingly, if the younger subjects are removed from the adult L2 group, that is, if we consider the older L2 adult group only, the percentage of targetlike responses in the scrambling condition no longer correlates with age at time of testing ($r = .132$, $p = .589$) or with length of exposure ($r = .306$, $p = .203$). The suggestion that it is the younger L2 adults who are probably responsible for the significant correlations observed in the original L2 adult group implies that due to the overlap in age at time of testing, these subjects pattern similarly to the L2 children, for whom similar correlations are also found.

6.4.3.2 Individual results

The results for individual subjects will now be examined. As with the group data, the individual data are first categorised according to proficiency level and subsequently according to age at time of testing.

6.4.3.2.1 Proficiency group

As with the negation task, the individual L2 subjects were categorised according to whether they had acceptance, rejection or mixed response patterns. Subjects who (incorrectly) rejected all or all but one of the test items were classified as having a pattern of rejection. Those who (correctly) accepted all or all but one of the test items were categorised as having a pattern of acceptance, and all subjects in-between were assigned to a 'mixed' response pattern.

In the no scrambling condition, except for the high proficiency children and the mid proficiency adults, where, respectively, 18.8% and 25% of subjects have a mixed or rejection pattern, the vast majority of subjects in each proficiency group for both the L2 children and L2 adults have the targetlike pattern of acceptance. As was the case for the L1 children and adults, it is only those subjects who sometimes or always reject the non-specific meaning for the scrambled form who also reject this meaning for the non-scrambled form (four children and four adults, 11.0% of all subjects). The subjects who unexpectedly reject the non-scrambled items appear to do so because they have rejected the scrambled items first. In other words, it appears that we are once again dealing with a carry-over effect.[25]

The results for the scrambling condition are given in Figure 7 for the L2 children and Figure 8 for the L2 adults. The exact numbers and individual data are provided in Appendix D. For both the L2 children and L2 adults, the

[25] Evidence that these subjects genuinely disallow the non-specific meaning for the non-scrambled form would come from a task which tested this latter order first and in which these subjects on that interpretation still rejected the non-scrambled items. For reasons outlined in §6.4.1.1, this order was not used here, but if the claim that these results are due to a carry-over effect is correct, then the prediction would be that we would no longer find subjects who reject the non-specific meaning for the non-scrambled form if this order were used instead.

proportion of subjects with the targetlike pattern of rejection increases with proficiency.

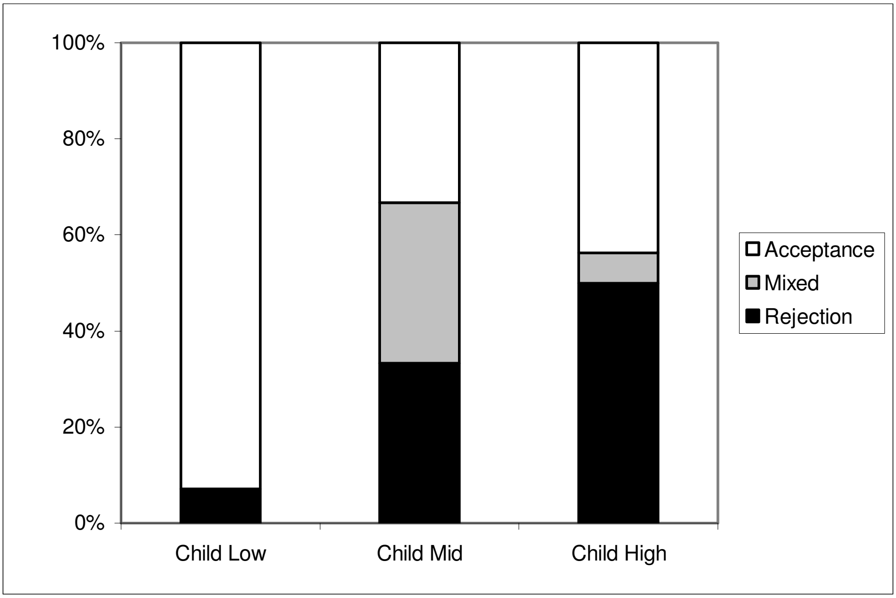


Figure 7. L2 children (comprehension): Distribution of individual response patterns (per proficiency group) for *twee keer* task (scrambling condition)

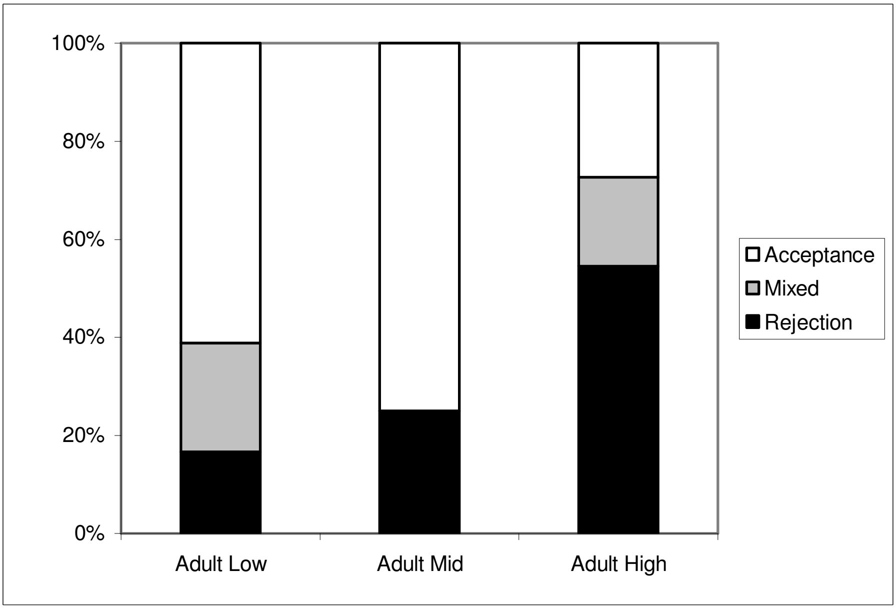


Figure 8. L2 adults (comprehension): Distribution of individual response patterns (per proficiency group) for *twee keer* task (scrambling condition)

Comparing responses on the two conditions for subjects other than those who produce the unexpected rejection response in the no scrambling condition, we find that nine children and 14 adults make a relative distinction between the two conditions, that is, they consistently accept the non-specific interpretation for the non-scrambled form and they sometimes or always reject this reading for the scrambled form. Amongst these subjects, there are seven children (one mid and six high) and eight adults (three low and five high) who make an absolute distinction between the two conditions.

6.4.3.2.2 Age at time of testing

In the no scrambling condition (see Appendix D for numbers), the majority of children in each group have a targetlike pattern of acceptance. As noted above, the exceptions, that is, those who sometimes or always reject the non-specific meaning for the non-scrambled form, are always subjects who provide targetlike responses in the scrambling condition; it was suggested that this unexpected non-targetlike behaviour in the non-scrambled condition, also observed in the L1 children and adults, results from a carry-over effect.

The individual response patterns for L2ers grouped by age at time of testing in the scrambling condition are given in Figure 9.

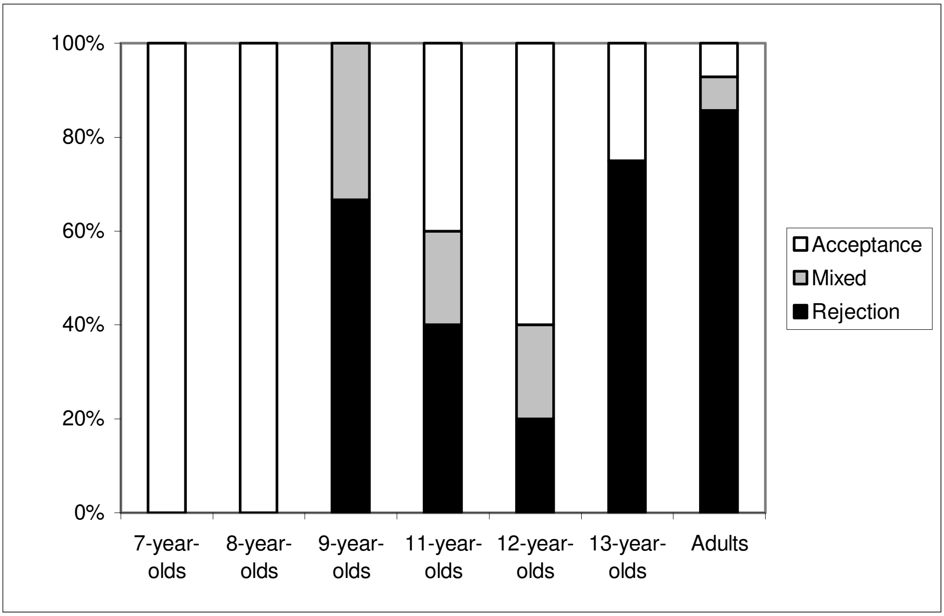


Figure 9. L2 children (comprehension): Distribution of individual response patterns (per age at time of testing) on *twee keer* task (scrambling condition)

All the children in the 7- and 8-year-old groups have a non-targetlike pattern of acceptance. In all remaining groups except the 12-year-olds, the majority of children have either a mixed or rejection pattern, that is, they correctly reject the non-specific interpretation for the scrambled form some or all of the time. (With respect to the 9-year-olds, it is perhaps worth bearing in mind that there are only three subjects in this group.) Amongst those subjects who do not exhibit a carry-over effect in the no scrambling condition, the subjects who make a relative or absolute distinction between the two conditions are distributed across the older age groups: two 9-year-olds, one 10-year-old, two 11-year-olds and five 13/14-year-olds make a relative distinction and one 9-year-old, one 11-year-old and five 13/14-year-olds make an absolute distinction.

6.4.3.3 Discussion

The L2 results in the judgement task on *twee keer* clearly differ from those on the negation task. Targetlike judgements are made by both L2 children and adults and the proportion of subjects providing such judgements increases with proficiency. There is also suggestive evidence of an age-effect in the child L2 data. The issues of proficiency and age are dealt with in turn. In addition, I consider whether the L2 results are compatible with Philip's (2003) and Krämer's (2000) approaches to the acquisition of scrambling. Finally, the predictions given in §6.1 are evaluated.

The results presented here show that it is possible for English-speaking L2ers to overcome the poverty-of-the-stimulus problem. There are both L2 children (11/36) and L2 adults (11/37) who consistently reject the non-specific reading for the scrambled indefinite object, and furthermore, the comments these subjects made during the experiment show that they reject the test sentences for the right reasons. Most of these subjects distinguish scrambled from non-scrambled objects, consistently accepting the non-specific reading for the non-scrambled form, and for those who do not, it was suggested that, like the L1 children and the one native adult who provided similar (unexpected) responses, this is due to a carry-over effect.

A proficiency effect is observed in the data. Although there is no significant correlation between scrambling responses and proficiency for the L2 adults, both L2 children and L2 adults generally become more targetlike with increasing proficiency. However, whereas the high proficiency L2 adults are not significantly different from the native adults, the high proficiency L2 children are, even though these two L2 groups do not significantly differ from each other in terms of proficiency (cf. Table 39). The pattern for the L2 adults becomes even clearer when this group is separated into younger and older subjects, that is, when L2 subjects whose age at time of testing is comparable to the L1 children are separated from those who are not. The difference between the high proficiency younger L2 adults and the adult natives is approaching significance, whereas the difference between the high proficiency older L2 adults and the natives is virtually non-existent. These findings contrast with those reported for the negation task, where no proficiency effect was observed. In §6.3.3.3, it was suggested that the absence of any proficiency effect might result from the fact that the proficiency measure was based on production data and the experimental task on comprehension data. The observation of a proficiency effect on the comprehension task with *twee keer* means that this difference in mode is unlikely to be responsible for the lack of such an effect on the negation task. In §6.4.3.3, I suggested that this mode difference might be an issue for some five subjects, that is, their low proficiency score might underestimate their L2 competence. On the *twee keer* task, three of these five subjects had a mixed response pattern and on the negation task, two of these had a targetlike pattern of acceptance. This is compatible with the claim that – assuming that targetlike knowledge of scrambling develops with increasing proficiency – these subjects are in fact more proficient than their (lack of) proficiency scores suggest.[26]

Let us now turn to the predictions made for L2 acquisition by Philip (2003). Recall that Philip claims that English-speaking L2ers are expected to take at least as long as L1 Dutch children. The L1 data presented in the preceding section suggest that it takes on average around nine years before the majority of Dutch children consistently produce targetlike responses on this task. This means that, on Philip's account, we should not find (a significant number of) targetlike English-speaking L2ers with less than nine years' exposure to Dutch.[27] The qualifier 'a significant number of' is added because finding one or two English-speaking L2ers with less than nine years of exposure who are targetlike would not necessarily falsify Philip's prediction for the reason that there are also a number of L1 children who are targetlike with less than nine years' exposure, that is, who are younger than nine years old.

Out of the 22 (child and adult) L2 subjects with a targetlike pattern of rejection in the scrambling condition, just seven (four children and three adults) have been exposed to Dutch for nine years or more, with length of exposure ranging from 9;1 to 16 years. In other words, the remaining 15 subjects (seven children and eight adults) have been exposed to Dutch for a period of less than nine years, ranging from 0;4 to 8;2. This constitutes counter-evidence to the prediction made on the basis of Philip (2003) that L1 children, L2 children and L2 adults should take as long as each other to acquire the specific interpretation of scrambled indefinite objects. It might be objected, however, that given that there are a handful of child L1 subjects younger than nine who provide targetlike responses, the 9-year criterion is not entirely fair.

[26] Note that I am not claiming that *because* these subjects provide targetlike responses (some or all of the time), they must be more proficient (as this would be circular).

[27] For present purposes, it will be assumed (counterfactually) that one year of exposure for an L1 child is the same as one year of exposure for an L2 child or adult. This is unlikely to be the case, however, given that both the quality and the quantity of the input to these two groups will be different. Given that all the L2ers tested here maintain use of their L1, the amount of exposure they receive *must* be less than the average L1 child.

Let us thus now consider whether there are subjects who have been exposed to Dutch for seven years or less who provide targetlike responses: 13 of the 22 targetlike subjects fit this criterion (five children and eight adults), with length of exposure ranging from 0;4 to 5;4. Even if we lower the criterion to four years, the youngest age at which (a very limited number of) targetlike L1 children have been found (Krämer 2000), there are still five targetlike L2 subjects (one child and four adults) with less than four years' exposure. If acquisition on the basis of indirect negative evidence is a long and gradual process, as Philip claims, these findings suggest that when evaluated in terms of the amount of exposure required for learners to demonstrate targetlike knowledge, such an account faces serious problems with the data presented here.

When the child L2 data are considered in terms of age at time of testing, it is only the younger subjects, the 7- and 8-year-olds, who are significantly different from the native adults. The difference between the youngest two groups and the 9-year-olds is also approaching significance. This is a similar pattern to that observed in the L1 data, where the 7- and 8-year-olds were also found to differ from the adults, and the difference between the 8-year-olds and the 9-year-olds was also approaching significance. In the child L2 data, there is a confound between age at time of testing, proficiency and length of exposure, however, and this makes interpreting the child L2 data rather difficult, as it is possible that the older L2 children's more targetlike responses may be due to their higher proficiency or their longer exposure to Dutch rather than to their age. The regression analysis carried out using all three variables found length of exposure to be the best predictor of targetlike responses in the scrambling condition; when the analysis included the first two variables only, however, it was age at time of testing which is the best predictor. In both these analyses, the predictor variables only accounted for relatively little of the variance in the data (ca. 25%), though. In other words, other factors must contribute to the observed variance.

A comparison with the older adult L2 data might help shed some light on this issue, as in this group, there was no effect of age at time of testing. Amongst the eight high proficiency older L2 adults, seven score 60% or higher, that is, they provide targetlike responses for three or more of the five items. This suggests that when age at time of testing is not a significant factor, high proficiency subjects can almost always be expected to be targetlike. If age at time of testing is a significant factor, as the above analysis suggests for child L2 acquisition, high proficiency L2 children younger than nine at time of testing should be non-targetlike. This is indeed the case: the high proficiency subjects who are seven or eight years old (subjects C20C, C9C and C25C) all fail to provide a single targetlike response. At the other end of the spectrum, half of the low proficiency older L2 adults (3/6) provide targetlike responses some or all of time (one scores 60% and the other two 100%).[28] If age at time of testing is a significant factor in child L2 acquisition, we would expect that if there are any targetlike children in the low proficiency child L2 group, these will be nine years or older. There is only one targetlike subject in the low proficiency child L2 group and he is nearly ten years old (9;10.9). The available data are thus compatible with the claim that age at time of testing is a crucial factor in the child L2 acquisition of the interpretive constraints on scrambled indefinite objects and, consequently, that limited discourse integration might play a similar role in child L2 acquisition as in child L1 acquisition. Further evidence in favour of this claim comes from the observation that the high proficiency older L2 adults do not differ from the natives, whereas, like the L2 children, the younger L2 adults, where age at time of testing may also play a role, do. In general, however, the relevant data are rather limited and, hence, in order to confirm this claim, more data would be needed.

The confound between proficiency and age at time of testing makes the predictions based on Krämer (2000) rather difficult to assess. The correlations between targetlike scrambling responses and proficiency, on the one hand, and between targetlike scrambling responses and age at time of testing, on the other, suggest that both L1 transfer and limited discourse integration play a role in the child L2 acquisition of the interpretive constraints on scrambling. In §6.1, it was suggested that if this were the case, it should be the most proficient of the older subjects who are targetlike. For the limited data available, this prediction was borne out. Furthermore, it was shown that L2 children younger than nine of an equivalent proficiency level to children aged nine or older were not targetlike. Further evidence of an age effect in the child L2 data comes from the observation that although an effect of proficiency was observed in both the L2 child and L2 adult groups, the most proficient L2 children remain significantly different from the natives, whereas the older L2 adults did not. This latter observation confirms prediction ❷-a, which stated that the younger L2 children should differ from the L2 adults. Combined with the observations made above, it also lends support to prediction ❹, which stated that these L2 children should pattern like L1 children. Prediction ❷-b stated that the older L2 children should pattern as the L2 adults and specifically it was predicted that if an effect of proficiency was observed in the latter, it should also be found in the former. To the extent that the relevant data were available, this prediction was also confirmed: those 13/14-year-olds who are non-targetlike were lower in proficiency than most of the targetlike children in this group. The relevant data were limited, however, and hence further research is necessary to explore this aspect of the data in more detail.

---

[28] This does not include the three low proficiency subjects discussed above and in §6.3.3.3, whose inability to complete the proficiency task might underestimate their proficiency level. These subjects belong to the younger L2 adult group.

6.4.4 Summary and conclusion

This section reported data on the comprehension of indefinite object NPs scrambled across the adverb *twee keer*. The results from the L1 children indicated that as with scrambling across negation, development is gradual and late. Adultlike levels of targetlike responses were observed at age 12. These results are compatible with Krämer's findings using an act-out task, but they contrast with Philip (2003), who observed much lower levels of targetlike responses, even at this late age. A difference in the complexity of the experimental scenarios was offered as a possible explanation for this discrepancy. The L1 results were fully consistent with Krämer's Non-Integration Hypothesis, which states that limited discourse integration is responsible for children's failure to assign the specific reading to the scrambled indefinite object.

In the L2 data, the proportion of targetlike responses was generally observed to increase with proficiency, across both child L2 and adult L2 groups. Unlike the L2 adults, however, the high proficiency L2 children remained significantly different from native adults. It was suggested that this difference stems from the effect of limited discourse integration which manifests itself in targetlike comprehension of scrambling being related to age at time of testing. Although the analysis of data was complicated by a confound with several variables, the limited data available were found to be compatible with this claim.

A comparison of these data with those of the previous section, that is, from the negation task, will be made in §6.6.1.

6.5 Task III: Act-out task with *twee keer*

A subset of the L2 adults carried out a third task, an act-out task, taken from Krämer (2000), which tested the contrast between the two sentences given in (22).

(22)	a.	Je	moet	een knikker	twee keer	rollen
		you	should	a marble	two times	roll
		'You should roll a (certain) marble twice.'				
	b.	Je	moet	twee keer	een knikker	rollen
		you	should	two times	a marble	roll
		'You should roll a(ny) marble twice.'				

On a targetlike reading, the subject is expected to roll one marble in (22)-a and one or two marbles in (22)-b. Recall that the non-scrambled sentence in (22)-b refers to two marble-rolling events and thus the number of marbles rolled is irrelevant. Krämer's results indicate, however, that despite having the option of rolling just one marble when presented with the non-scrambled sentence, native adults and children (virtually) always choose to roll two different ones.

The purpose of carrying out this task with the L2ers was to determine whether they assign a targetlike specific interpretation to the scrambled indefinite object and in particular whether they demonstrate the same preference as adult (and child) native speakers.

Recall that the English equivalent of the scrambled or non-scrambled sentences given in (22) is ambiguous between a specific and non-specific meaning:

(23) You should roll a marble twice

In response to (23), an adult native-speaker of English could manipulate either one or two marbles. In order to determine whether this ambiguity is found in a task situation, a pilot study was carried out with English native adults. The task employed the same format as the one described in the following section for Dutch. It included five test items plus fillers. Ten adults were tested. The results indicate that when instructed to carry out an action as in (23), native English adults prefer to manipulate one object rather than two. The average number of 1-object responses across subjects for such sentences was 78%. Most (7/10) of the subjects consistently manipulated one object only; one consistently manipulated two objects, and the remaining two subjects sometimes manipulated one and sometimes two objects. These results mean that if an English-speaking L2er consistently manipulates one object when presented with a scrambled order in Dutch, as in (22)-a, this could result from either targetlike knowledge of the interpretive constraints on scrambling or L1 transfer. In order to distinguish between these two options, it is necessary to consider the L2ers' responses to non-scrambled orders, as in (22)-b, as well. If in response to this order, L2ers manipulate two objects, they can be said to make the targetlike distinction.[29]

6.5.1 Method

The method replicated Krämer's task with a number of modifications. Subjects were asked to carry out a series of tasks, such as the one given in (22), using a number of props laid out on a large table. Given that one of the aims of this task is to determine whether L2ers make the relevant contrast between the scrambled and non-scrambled sentences, it was necessary to use a within-subjects design, that is, as in the truth value judgement task with *twee keer*, each subject was tested in a scrambling and a no scrambling condition (cf. Krämer's task which employed a between-subjects design – see §3.3.1.1 for relevant discussion).

[29] Manipulating one object only in response to the non-scrambled order is uninformative as it is consistent both with L1 transfer and the dispreferred (yet nevertheless grammatical) target language response.

6.5.1.1 Materials

There were five test items per condition and for each item, there were four objects, distinguished by colour or size. In the scrambling condition, exactly the same predicates as in Krämer's original experiment were used. The five items in the no scrambling condition used the same verbs (*rollen* 'to roll', *omdraaien* 'to turn over', *wegblazen* 'to blow away', *indrukken* 'to press' and *gooien* 'to throw') but with different objects.[30] A complete list of test and filler items are given in Appendix D. The test items were interspersed with 16 fillers, which were designed to distract and check that the subject was paying attention. The fillers were of three types: (i) the original five fillers from Krämer's experiment, (ii) fillers which required moving two objects rather than carrying out two actions, and (iii) fillers which asked for an action to be carried out three (rather than two) times. The latter type of items contained a definite NP object which was either scrambled or not, depending on the condition it was used in. Examples of all three types of filler are given in (24).

(24) a. Je moet de blauwe knikkers in een kopje doen
you should the blue marbles in a cup do
'You should put the blue marbles in a cup.'

b. Je moet twee vogels over de boom laten vliegen
you should two birds over the tree let fly
'You should make two birds fly over the tree.'

c-i. Je moet het schaap drie keer over een tijger laten springen
you should the sheep three times over a tiger let jump
'You should make the sheep jump over the tiger three times.'

c-ii. Je moet drie keer het schaap over een tijger laten springen
you should three times the sheep over a tiger let jump
'You should make the sheep jump over the tiger three times.'

As in the previous tasks, the intonation used for all items was the most natural for the scrambled order, namely with the pitch rise for the VP starting on *twee keer* 'twice' and the indefinite determiner was always pronounced as *een* \ən\ rather than *één*. This was the same for the non-scrambled items.

6.5.1.2 Procedure

The act-out task followed the truth value judgement task on negation (see §6.3.1.2 for the reasoning behind this ordering). Subjects were told that they would be asked to carry out a series of tasks using the props laid out on a large

---

[30] The new items were trialed in the English pilot study mentioned above. Some items had to be redesigned as they induced a response bias. For example, one item with the verb 'to press' required subjects to press a button which released a toy car from its holder. The fact that subjects had to put the car back in its holder before they could press the button again meant that subjects tended to manipulate two objects rather than one.

table in the testing room. The experimenter first took the subject through each set of objects, explaining what they were called and demonstrating how they should be used. In particular, she showed the subjects how, after using an object, they should place it back in its original position before selecting it or another object. This instruction was added to discourage subjects from keeping one object in their hand and using this to carry out all actions. This was particularly important given the native English adults' preference for a specific interpretation in the English pilot study mentioned above. In her explanation, the experimenter always manipulated two different objects and one of these was manipulated twice, so that the subject could see that this was allowed. Subjects were instructed to place any objects they had manipulated on a paper napkin which was placed next to each set of objects. This meant that the experimenter could clearly see how many objects the subjects had used when it came to coding their responses. The test items were preceded by two warm-up items, which allowed the subjects to familiarise themselves with the task. As in the truth value judgement task on *twee keer*, a blocked design was employed. This time, however, the non-scrambled items preceded the scrambled items. This was because in a pilot study, it was found that native Dutch adults are more likely to manipulate one object in the non-scrambled condition, which, recall is perfectly legitimate, if they have heard the scrambled condition first, whereas in the reverse order, that is, with the non-scrambled condition preceding the scrambling condition, they consistently manipulate two objects in the non-scrambled condition.

6.5.2 Subjects

Fourteen of the L2 adults also took part in the act-out task.[31] Their age at time of testing ranged from 22;2 to 50;0 (mean=34;8; SD = 9;5), their age at first exposure from 21;10 to 35;0 (mean=27;1; SD = 5;8), their length of exposure from 0;4 to 25;0 (mean=7;5; SD = 6;10) and their proficiency score from –1.26 to 1.37 (mean=.23; SD = 0.91). Individual data are given in Appendix D. All these adults were adults at time of testing.

6.5.3 Results

The results are first presented for the group as a whole and subsequently for individual subjects.

6.5.3.1 Group results

Subjects' responses were categorised according to whether they manipulated one or two objects. Table 56 presents the average percentage of 1-object responses for the L2 proficiency groups and the native adults.

---

[31] I had intended to test all L2 adults on the act-out task, but for practical reasons, this was not always possible. Space and time constraints also meant that it was unfortunately not possible to test the L2 children on the act-out task.

Table 56. L2 adults (comprehension): Group results (organised by proficiency) on act-out task (both conditions). Average percentage of 1-object responses.

	n	No scrambling condition		Scrambling condition		Mann-Whitney tests comparing different conditions	
		mean	SD	mean	SD	Z	Significance
Low	5	64.0	43.4	84.0	35.8	-1.633	p = .102
Mid	2	60.0	28.3	60.0	56.6	.000	p = 1.000
High	7	49.6	39.8	84.3	9.8	-2.041	p = .041
Controls	14	14.2	30.9	88.6	21.8	-3.241	p = .001

The native adults make the expected distinction: they generally produce 1-object responses in the scrambling condition and 2-object responses in the no scrambling condition. The handful of unexpected 2-object responses (n=8) in the scrambling condition could be the result of a carry-over effect from the no scrambling condition which was presented first, although the 88.6% rate of 1-object responses in this condition does not differ substantially from the 92% produced by Krämer's adult controls, who were tested on the scrambling condition only. Although the native adults tested here also produce a number of 1-object responses (n=10) in the no scrambling condition (cf. Krämer's controls, who produced none), they still make a clear, significant distinction between the two condition, as the results of the Mann-Whitney test in the final column of Table 56 indicate.

Turning to the L2 adults, all three groups produced on average more 1-object than 2-object responses in the scrambling condition. This pattern is replicated in the no scrambling condition for the low and mid proficiency groups, and the high proficiency group produce on average 1-object responses in half the items. Whether this is the result of chance behaviour will be considered in the following section where the individual results are presented. Like the native adults, for high proficiency L2 adults, there is also a significant difference in the number of 1-object responses produced in the scrambling vs. no scrambling conditions.

6.5.3.2 Individual results

Subjects who provided 1-object responses in all or all but one of the test items were classified as having a 1-object response pattern. Those who provided 2-object responses in all or all but one of the test items were classed as having a 2-object response pattern, and anything in-between was characterised as a mixed response pattern. The results of this classification are provided in Figure 10 for the scrambling condition and Figure 11 for the no scrambling condition. The no scrambling condition is presented second here because the import of the results in that condition is best understood after examining the results in the scrambling condition. The black shaded areas are the target (preference) responses, that is, a 1-object response in the scrambling condition and a 2-object response in the no scrambling condition.

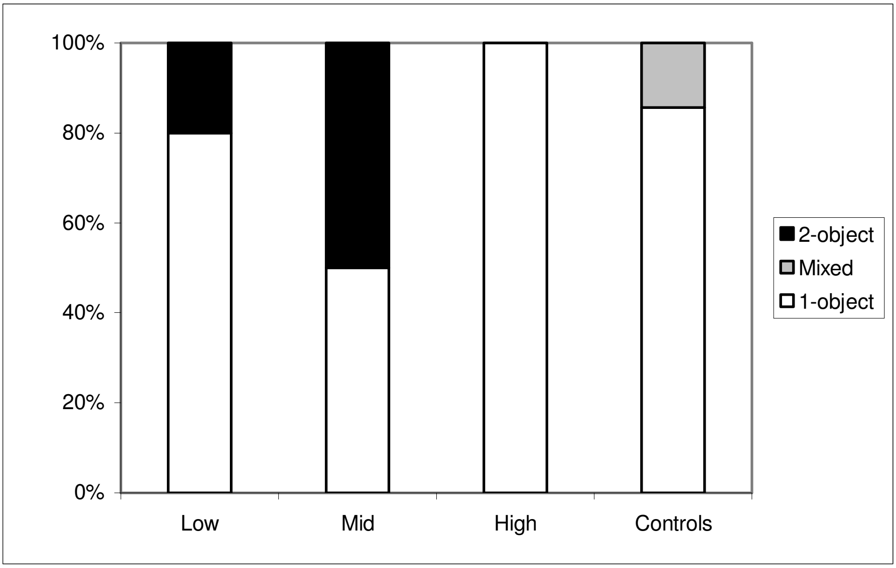


Figure 10. L2 adults (comprehension): Distribution of individual response patterns (per proficiency group) on act-out task (scrambling condition)

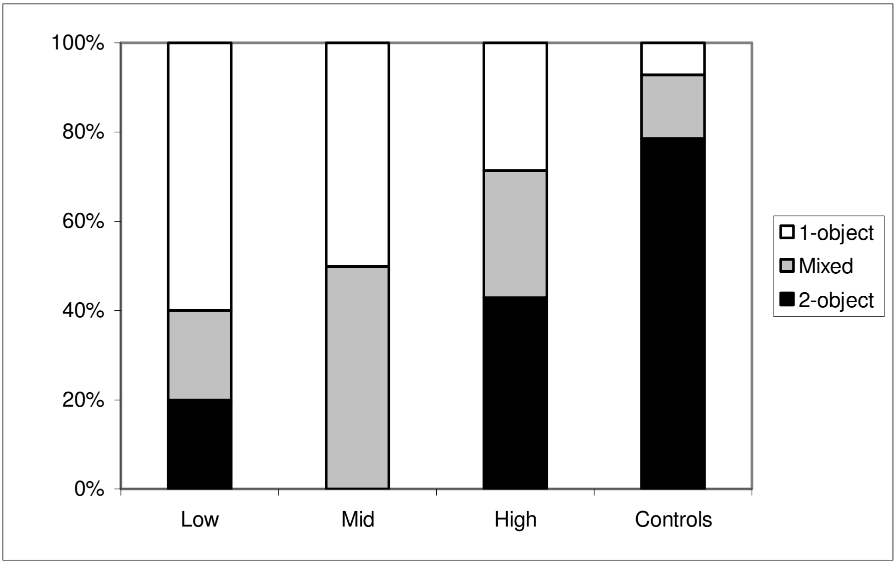


Figure 11. L2 adults (comprehension): Distribution of individual response patterns (per proficiency group) on act-out task (no scrambling condition)

In the scrambling condition (Figure 10), the vast majority of subjects in each group (with the exception of the mid proficiency group, where there are only two subjects) have a 1-object response. This could be the result of knowledge of the specific meaning for the scrambled order or it could be transfer from L1 English. The same holds for a 1-object response in the no scrambling condition; this could also be the result of L1 transfer. The only truly informative responses are 2-object or mixed response patterns in the no scrambling condition combined with a 1-object response pattern in the scrambling condition. This would provide evidence that an interpretive distinction exists between the scrambled and non-scrambled orders. Most (85.7%) of the native adults have this combination of response patterns, as do six of the 14 L2 adults. Of these six, one is in the low proficiency group and five are in the high proficiency group.[32] How these subjects fare on the truth value judgement task on *twee keer* is discussed in §6.6.2.

6.5.4 Discussion

The act-out task was employed as a supplementary task to determine whether L2 adults make a targetlike distinction between scrambled and non-scrambled indefinite objects and, in particular, whether they exhibit the same preferences as native speakers. Upon hearing a sentence with a scrambled object, the vast majority of L2 subjects provided a 1-object response. This could reflect knowledge of the interpretive constraints on scrambling, that is, it could be a targetlike specific interpretation of the scrambled objects or it could be the result of transfer from L1 English, where for the translational equivalent, the specific reading is preferred. A comparison of L2ers' responses in the scrambling condition with those in the no scrambling condition revealed that more than half of the L2ers make a targetlike distinction between the two, that is, they consistently manipulated one object in the scrambling condition, and in the no scrambling condition, they manipulate two objects some or all of the time. (Note again that those subjects who manipulate one object only in both conditions may know the constraints on scrambling but because a 1-object response in the no scrambling condition is not incorrect, their responses are uninformative – see §6.6.2 for further discussion.) This result suggests that not only do these L2ers have the targetlike specific reading for the scrambled object, they also have the same preferences regarding the interpretation of the non-scrambled object.

The L1 children tested by Krämer mainly have 2-object responses in the no scrambling condition from the youngest group (4 years old) onwards. This contrasts with the L2 adults, who predominantly show a 1-object response pattern even at the low proficiency level. This difference lends support to the claim that the preference for a 1-object response is the result of

[32] The number of subjects per cell is too few to carry out a Chi-squared analysis to determine whether the distribution of response patterns differs between groups.

L1 transfer. In the scrambling condition, the L1 children move from generally having 2-object responses to having more 1-object responses. Again, this contrasts with the L2 adults, who, because they start with 1-object responses as a result of L1 transfer, do not show any development in the scrambling condition. The real indicator of targetlike knowledge is the distinction between the two conditions. This develops with increasing proficiency, that is, it is (almost exclusively) the high proficiency subjects who make the targetlike distinction.[33]

6.5.5 Summary and conclusion

This section considered whether L2 adults are able to come to know the interpretive constraints on scrambled indefinites by means of an act-out task. The data indicated that although their responses were not identical to the native controls, the high proficiency group consistently provided 1-object responses in the scrambling condition and for the majority, 2-object or mixed responses in the no scrambling condition. In other words, they demonstrated knowledge of the specific interpretation of the scrambled indefinite object and of the native preference for a 2-object response in the no scrambling condition.

6.6 Cross-task comparisons

This section compares the results of the three experiments with each other. First, the two truth value judgement tasks are compared and subsequently the *twee keer* judgement task with the act-out task.

6.6.1 Comparison of two truth value judgement tasks

Two truth value judgement tasks testing scrambling, one with negation and the other with *twee keer*, were carried out with all three learner groups. This means that it is possible to compare responses on these two tasks within the same child (cf. Krämer, who used a between-subjects design). This will allow us to determine whether scrambling across negation and scrambling across *twee keer* develop in parallel or whether targetlike comprehension of one is in place before the other. Responses on these two tasks will first be compared within each of the three groups, in §6.6.1.1, and subsequently, in §6.6.1.2, the findings for each of these groups will be compared with each other. We will see that for

[33] Note that the native English adults' preference for a 1-object response in the pilot English act-out task does not undermine the claim that the acquisition of the interpretive constraints on scrambled indefinites in Dutch presents a poverty-of-the-stimulus problem for the English-speaking L2er. The poverty-of-the-stimulus problem concerns how these L2ers rule out the non-specific (i.e. 2-object) reading for the scrambled indefinite; this is tested in the truth value judgement task, where subjects are asked to do just that. Even though English native speakers generally have a preference for a 1-object manipulation in the act-out task, they would not rule out a 2-object manipulation as impossible.

all three learner groups, there is no clear relation between responses on the two tasks. Although most subjects provide either targetlike or non-targetlike responses on both, there is a not insignificant number of learners who are either targetlike on the negation task and non-targetlike on the *twee keer* task or vice versa. In §6.6.1.3, possible explanations for these findings are considered. These include certain unavoidable differences between the two tasks, such as the targetlike response ('no' in the *twee keer* task and 'yes' in the negation task), as well as the potentially infelicitous use of negation in that task. I argue that this, combined with L1 transfer, may explain the generally non-targetlike response patterns observed in the L2 data.

6.6.1.1 Within-group comparisons

Each set of learners' responses will be considered as a group and individually.

6.6.1.1.1 L1 acquisition

Wilcoxon signed-ranks tests were carried out to determine whether, for the scrambling conditions, the percentage of targetlike responses in the negation experiment differs from the percentage of targetlike responses in the *twee keer* experiment.[34] The results, given in Table 57, shows that for most of the child L1 groups and for the adult controls, no distinction is made between the two.[35] For the 12- and 13-year-olds, however, there is a significant difference between responses on the negation experiment and responses on the *twee keer* experiment. This is because these children are targetlike on *twee keer*, but not on negation, whereas the younger groups are equally non-targetlike on both (and the adults are equally targetlike on both).

Table 57. L1 children and adults (comprehension): Negation compared with *twee keer*. Average percentage of targetlike responses in scrambling condition

	n	Negation (target: yes)		*Twee keer* (target: no)		Wilcoxon signed-ranks test	
		%	SD	%	SD	Z	Significance
7-year-olds	14	10.0	25.7	25.7	41.8	-1.370	p = .171
8-year-olds	15	20.0	37.0	18.7	38.9	.000	p = 1.000
9-year-olds	15	65.3	41.7	56.0	40.1	-.953	p = .341
11-year-olds	15	57.3	44.0	60.0	46.0	-.060	p = .952
12-year-olds	13	75.4	29.6	93.9	22.2	-1.983	p = .047
13-year-olds	25	77.0	33.8	98.0	10.0	-2.803	p = .005
Adults	13	92.9	14.9	90.0	28.0	-.137	p = .891

[34] Wilcoxon signed-ranks is a non-parametric test which can be used when the data do not fulfil the requirements for a paired t-test. All results reported here are based on positive ranks.
[35] The number of subjects per group differs from those given in §6.3.2 and §6.4.2 as there were three subjects who were excluded from one of the two experiments. Table 57 only includes subjects who completed both.

There is a significant moderate correlation between children's responses on the two tasks (r = .558, p = .000). Combined with the group results in Table 57, this suggests that for the younger groups at least, subjects' responses on the two tasks are more or less the same. In order to ascertain whether this really is the case, it is necessary to examine the results for each subject individually (see Appendix D for individual data). The following broad patterns can be discerned. There are:

(a) subjects who have a mixed or targetlike response pattern on the negation task and on the *twee keer* task (n=50);

(b) subjects who have a mixed or targetlike response pattern on the negation task but a non-targetlike response pattern on the *twee keer* task (n=9);

(c) subjects who have a mixed or targetlike response pattern on the *twee keer* task but a non-targetlike response pattern on the negation task (n=14);

(d) subjects who have a non-targetlike response pattern on both tasks (n=24).

Amongst the 50 subjects with the first pattern, 43 have a targetlike response pattern on both tasks. As expected, the number of children with this pattern increases with age: there are two out of 14 (14.3%) 7-year-olds; one out of 15 (6.7%) 8-year-olds; five out of 15 (33.3%) 9-year-olds and 11-year-olds; nine out of 13 (69.2%) 12-year-olds and 21 out of 25 (84%) 13-year-olds. This increase is accompanied by a concomitant decrease in the number of children who are non-targetlike on both tasks (pattern (d)): eight out of 14 (57.1%) 7-year-olds; nine out of 15 (60%) 8-year-olds; three out of 15 (20%) 9-year-olds and 11-year-olds; and no 12- or 13-year-olds. There are a number of children (23.7% (23/97)) who are more targetlike on one task than on the other. These are more or less equally split across both tasks, that is, there are around the same number of children who are more targetlike on the negation task than on the *twee keer* task (pattern (b)) as children who are more targetlike on the *twee keer* task than on the negation task (pattern (c)). Neither of these two patterns is specific to a particular age group, that is, both patterns are relatively evenly distributed across the different age groups. Note, however, that in the oldest group, there are only subjects who are better on *twee keer* than on negation.[36] It seems, thus, that for around one fifth of the child L1 subjects, their responses

[36] The data were analysed in a number of alternative ways, including comparing the number of targetlike responses on each task relative to each other (rather than comparing response patterns) and splitting the mixed pattern such that responses of 60% (3/5) or more were contrasted with those less than 60%. These analyses resulted in slightly different numbers of subjects per category, but the overall pattern remained more or less the same and, importantly, there were always some subjects who were more targetlike on the negation task than on *twee keer* and vice versa.

on the two tasks do not necessarily match each other. How this finding should be interpreted will be discussed in §6.6.1.3.

6.6.1.1.2 L2 acquisition

A similar analysis was carried out with the L2 data. The group data and results of the Wilcoxon signed-ranks tests are given in Table 58. There was no significant correlation between subjects responses on the two conditions for either group (L2 children: r = -.040, p = .821; L2 adults; r = .118, p = .488).

Table 58. L2 children and adults (comprehension): Negation compared with *twee keer*. Average percentage of targetlike responses in scrambling condition

	n	Negation (target: yes)		*Twee keer* (target: no)		Wilcoxon signed-ranks test	
		%	SD	%	SD	Z	Significance
L2 children							
Low	12	35.0	.46	7.1	.27	-1.294	p = .196
Mid	6	50.0	.41	53.3	.45	-.105	p = .916
High	16	26.3	.40	50.0	.47	-1.557	p = .120
L2 adults							
Low	18	35.6	.44	31.1	.38	.461	p = .645
Mid	8	2.5	.07	25.0	.41	-1.289	p = .197
High	11	40.0	.46	63.6	.44	-1.015	p = .310

Responses on the two tasks are roughly the same for the mid proficiency children and the low proficiency adults but for the remaining adult groups and the high proficiency child group, scores on the *twee keer* task are higher than those on the negation task. The low proficiency children show the reverse pattern.

As with the L1 children, the results were examined for each subject individually (see Appendix D for data). There are:

(a) subjects who have a mixed or targetlike response pattern on the negation task and on the *twee keer* task (n=15: L2 children = 7, L2 adults = 8);

(b) subjects who have a mixed or targetlike response pattern on the negation task but a non-targetlike response pattern on the *twee keer* task (n=12: L2 children = 7, L2 adults = 5);

(c) subjects who have a mixed or targetlike response pattern on the *twee keer* task but a non-targetlike response pattern on the negation task (n=16: L2 children = 7, L2 adults = 9);

(d) subjects who have a non-targetlike response pattern on both tasks (n=28: L2 children = 13, L2 adults = 15).

Amongst the 15 subjects with the first pattern, four have a targetlike pattern on both tasks (C21C (high), C33C (high), A32C (low), A24C (high)). The remaining 11 subjects are two mid and three high proficiency children and four low and two high proficiency adults. The children who have a non-targetlike response pattern on both tasks (pattern (d)) are split between the low (n=7) and high (n=6) proficiency groups, but the adults are predominantly from the low (n=8) and mid (n=6) groups (cf. high: n=1). There are roughly the same number of subjects who are more targetlike on the negation task than on the *twee keer* task as those with the opposite pattern. These subjects are split relatively evenly across the child and adult groups and across the three proficiency levels within these groups. Subjects with pattern (c), that is, those who are better on *twee keer* than negation, predominate in the high proficiency groups, however: compare five out of 11 (45.5%) adults and four out of 16 (25%) children with pattern (c) vs. pattern (b) with two out of 11 (18.2%) adults and one out of 16 children (6.3%). The predominance of pattern (c) is comparable to the older L1 children.

The child L2 data were also analysed on the basis of age at time of testing, as in Table 59. The difference between the two tasks is almost significant for the 8-year-olds, who provided some targetlike responses on negation but none on *twee keer*, and it is likewise approaching significance for the 13/14-year-olds, but they are more targetlike on *twee keer* than on negation. The 9-year-olds are also more targetlike on *twee keer* than on negation, and the remaining three groups have more or less similar patterns on both tasks.

Table 59. L2 children (comprehension): Negation compared with *twee keer*. Average percentage of targetlike responses in scrambling condition

Age group	n	Negation (target: yes)		*Twee keer* (target: no)		Wilcoxon signed-ranks test	
		mean %	SD	mean %	SD	Z	Significance
7-year-olds	3	0	0	0	0	.000	p = 1.000
8-year-olds	10	38.0	49.4	0	0	-1.890	p = .059
9-year-olds	3	13.3	23.1	86.7	23.1	-1.633	p = .102
10-year-olds	5	48.0	50.2	52.0	50.2	.000	p = 1.000
11-year-olds	5	48.0	50.2	32.0	46.0	-.743	p = .458
13/14-year-olds	8	30.0	33.8	67.5	42.7	-1.876	p = .061

Patterns (a) through (d) observed above for the individual data are distributed across the different child L2 age groups as follows: nine of the 13 7-year-old and nine of the 13 8-year-olds have non-targetlike patterns on both tasks and the remainder are non-targetlike on *twee keer* and targetlike on negation. For the 10- and 11-year-olds, the remaining patterns are relatively evenly distributed; pattern (b) (negation > *twee keer*) does not occur in the 9- and 13-year-old groups.

6.6.1.2 Cross-group comparisons

In the preceding section, each group's results on the two truth value judgement tasks were compared with each other. In this section, the patterns observed in the child L2 data are first compared with those in the adult L2 data, and subsequently, with those in the child L1 data.

6.6.1.2.1 Child L2 ~ adult L2

The rates of targetlike responses on the two truth value judgement tasks do not differ from each other for any of the child L2 and adult L2 proficiency groups. There are, nevertheless, two differences between the L2 children and the L2 adults. First, the proportion of subjects who have a mixed or targetlike response pattern in both conditions (pattern (a)) increases with proficiency in the child L2 group (from 0% in the low group to 33.3% (2/6) in the mid group and a comparable 31.3% (5/16) in the high group), whereas this is not the case in the adult L2 group (cf. 27.8% (5/18) in the low, 0% in the mid and 27.2% (3/11) in the high group).[37] Second, the proportion of subjects who are non-targetlike on both tasks is relatively high in the high proficiency child L2 group (37.5% (6/16)) compared with the high proficiency adult L2 group (9.1% (1/11)).[38]

There are also two similarities between the L2 children and adults. First, there are only two subjects in each group who are completely targetlike on both tasks. This is largely due to the relatively low number of targetlike subjects on the negation task compared with the *twee keer* task. Second, in both groups we find subjects who are better on negation than on *twee keer* and subjects who have the reverse pattern. In other words, for both L2 children and L2 adults, targetlike responses on one task are not necessarily associated with targetlike responses on the other.

6.6.1.2.2 Child L2 ~ Child L1

The greatest difference between the L1 and L2 children (and between the L1 children and the L2 adults) is the proportion of subjects who are completely targetlike on both tasks. Out of the 93 L1 children, 41 (44.1%) have a targetlike response pattern on both tasks, whereas just four out of the 36 (11.1%) L2 children do so. Note, however, that this comparison is not entirely fair as the L2 sample includes subjects who may not be targetlike because of their age *and* subjects who may not be targetlike because of their low proficiency.

---

[37] Note however, that two of the five subjects in the adult L2 low proficiency group are subjects whose inability to complete the proficiency task might not be an accurate reflection of their proficiency level (see §6.3.3.3 and §6.4.3.3 for discussion), in which case they might belong to a higher proficiency level.

[38] The relative distribution of the L2 adults is more or less the same when the younger L2 adults are removed. The only major difference is that the number of low proficiency subjects who are non-targetlike on both tasks decreases (from 44.4% (8/18) to 16.7% (1/6)).

There are two points of similarity between the L1 and L2 children. First, the oldest groups, the L2 13/14-year-olds and the L1 12- and 13-year-olds, are both significantly better on the *twee keer* task than on the negation task. Secondly, there are both L1 and L2 children who are better on negation than on *twee keer* and vice versa. In other words, the L1 children are similar to the L2 children (and adults) in that targetlike responses on one task do not necessarily go hand in hand with or imply targetlike responses on the other task.

6.6.1.3 Discussion

In all three groups, the L1 children, the L2 children and the L2 adults, we find subjects whose responses on one task do not necessarily correspond with their responses on the other. In this section, I explore possible explanations for this finding. First, I consider whether scrambling across negation and scrambling across *twee keer* should be analysed as two different processes. Arguing that such a conclusion would be premature, I then go on to consider some methodological differences between the two tasks. Ultimately, I suggest that there may be a number of factors at play, including certain unavoidable differences between the two tasks and the potentially infelicitous use of negation in the negation task. In combination with L1 transfer, this is argued to lead to the generally poor performance of the L2 subjects on the negation task.

The first approach towards accounting for the observed differences between the two tasks would be to take the results at face value and to reason that given the observed differences between the two tasks, these tasks must be tapping different types of knowledge. Such a conclusion would lead us to ask whether, contrary to the analyses discussed in Chapter 3, scrambling across negation somehow differs from scrambling across *twee keer*. Recall that scrambling across negation and scrambling across *twee keer* are supposed to involve the same type of movement. Although negation interacts with focus in the scrambling of definite NP objects, scrambling of an indefinite object NP is claimed to induce the same type of semantic effects irrespective of the scope-taking operator across which it scrambles. To conclude on the basis of the current results that this is not the case would, I fear, be a little hasty. Before we can seriously entertain this possibility, the second explanation should be considered.

The second explanation for the observation that some subjects' responses on the negation task do not correspond with their responses on the *twee keer* task and vice versa is methodological. In a recent paper, Gualmini (2003) suggests that young children's problems with the specific reading of indefinites in negative sentences could (at least in part) result from the infelicitous use of negation in the experiments in question. He notes that negative sentences are easier to process in contexts where particular felicity conditions are met (De Villiers and Tager-Flusberg 1975; Wason 1965). For example, uttering the sentence 'It's not raining' would sound odd without the expectation that it should be raining. Similarly, Gualmini argues, failure to meet such an expectation in experimental scenarios will also mean that negative sentences are somewhat odd. In a truth value judgement task on L1 English, Gualmini presented children (n=30; 4;0-5;5) with one of two sentences, both of which had the same truth conditions but they differed with respect to their felicity. Thus, the sentence in (25)-a was felicitous in the context of the story in which it was presented because it expressed a mismatch between the expectation developed in the story (that the fire-fighter, who was playing hide-and-seek, would find all the dwarves) and the actual outcome of that story (that the fire-fighter only found some of the dwarves). The sentence in (25)-b, on the other hand, was infelicitous because it expressed no such mismatch, because the story in which it was presented did not lead to the expectation that the fire-fighter would miss all the dwarves.

(25) a. The fire-fighter didn't find some dwarves
b. The fire-fighter didn't miss some dwarves

Gualmini observes that this difference in felicity is indeed reflected in children's responses: sentences like (25)-a were accepted in 90% (54/60) of trials compared with an acceptance rate of just 50% (30/60) for trials involving sentences like (25)-b. Gualmini's point could be applicable to the experimental scenario used in the negation task here. Although at the beginning of the test story (see (3) in §6.3.1), the experimenter suggests that the main character wants to manipulate the objects, for example, that he wants to catch the fish, this is only a suggestion, and the story certainly does not build up any expectation that *all* the objects will be manipulated, as in Gualmini's experiment. It is therefore possible that the use of the negative test sentence was perceived as infelicitous by some of the (L1/L2) subjects and this may have contributed to their non-targetlike responses.

It is interesting to note that even the native adult controls in Gualmini's task made a difference between the felicitous and infelicitous sentences: four control subjects accepted sentences like (25)-a in 81.3% (13/16) trials, whereas 14 other controls accepted sentences like (25)-b in only 48.2% (27/56) of trials. Although this contrast is difficult to evaluate without data from individual subjects, other experimental research testing similar sentences, such as (27), suggests that English native-speaker adults vary in their willingness to assign a specific reading to indefinite object NPs embedded under negation and that their responses may vary between subjects and between trials (see discussion of Miller and Schmitt (2003) and Su (2001) in §3.1.1.3).

(26) Mickey didn't ride a dog

Such experiments on English differ from those on Dutch, however, because unlike the English sentences in (25) and (27), the scrambled test sentence used in the Dutch negation experiment is unambiguous. This means that even if Dutch adults were to perceive the use of negation in the experimental scenarios to be infelicitous, for this to be reflected in their responses, they would have to attribute an ungrammatical reading to the scrambled object. In other words, pragmatics would somehow have to 'overrule' the grammar. For the majority (12/14) of native adults tested here, this does not seem to have occurred, although there are two subjects who sometimes allow the (ungrammatical) non-specific reading for the scrambled indefinite (in two out of five items each). It is nevertheless plausible that a developing grammar might be affected by such pragmatic considerations, especially if, as in the case of the English-speaking L2ers tested here, test subjects have another source of knowledge, namely the L1, which allows the reading that is ungrammatical in the target language. What I am suggesting, then, is that the English-speaking L2ers may be more likely to allow the non-specific reading of the scrambled indefinite on the negation experiment than on the experiment with *twee keer* as a combined result of the potentially infelicitous use of negation in the experimental scenario and L1 transfer. This would account for the L2ers' generally poorer performance on the negation task. Ultimately, however, this remains an empirical question. If the suggestion made here is along the right lines, we would expect that if the use of negation in the experimental scenario were made more felicitous, as in Gualmini's experiment, more L2ers – and in particular those high proficiency L2ers who are targetlike on the *twee keer* task – would correctly accept the specific reading of the indefinite object scrambled over negation. With respect to the L1 children, we would predict an increased rate of acceptance, and we might expect the discrepancy between the results on the two tasks in the older children to be reduced considerably.

If there is a methodological problem with the negation experiment, this could account for the observation that there are some subjects who are targetlike on the *twee keer* task but not on the negation task. In each of the three learner groups, there was, however, a similar proportion of subjects who were successful on the negation task but not on the *twee keer* task. There are two differences between the tasks which might account for this finding. First, in the negation task, the object to which the scrambled indefinite refers is present in the pictures used in the experiment, whereas this is not the case in the *twee keer* task. Recall that on the latter task, the main character manipulates two different objects, an action which is consistent with the non-specific reading of the object. The specific reading must be *inferred* by the subject. This could potentially make the *twee keer* task more difficult.[39] The second difference

[39] It might be objected that the *twee keer* task does not conform to the condition of plausible dissent (Crain and Thornton 1998), and that this may contribute to the high proportion of non-targetlike responses, especially amongst the child subjects. This seems unlikely, however, given that the condition of plausible dissent was built into Philip's experimental scenario and his (child L1) subjects were generally more *non*-targetlike than those here.

concerns the targetlike response on each of the two tasks. Due to the different entailment relations between the scrambled and non-scrambled sentences, the experimental scenario used in the negation task verifies the targetlike reading of the test sentence, that is, the targetlike response is 'yes'. The scenario used in the *twee keer* task, however, falsifies the targetlike reading of the test sentence and consequently the targetlike response is 'no'. It is possible that this discrepancy may also be reflected in the differences between responses on the two tasks observed for some subjects. If this were the case, we would expect that on a task which tested the comprehension of scrambling across *twee keer* using an experimental scenario which verified the targetlike reading, subjects should produce comparable responses on this and the negation task (as long as any problems with the infelicity of negation, as discussed above, were first rectified of course).[40, 41]

6.6.1.4 Summary

A comparison of subjects' responses on the two truth value judgement tasks indicates that although the majority of subjects are either targetlike on both or non-targetlike on both, there is a not insignificant number of subjects whose responses on one task differ from their responses on the other. This holds for all three learner groups. It was suggested that rather than concluding that scrambling across negation and scrambling across *twee keer* must be different from each other, a number of methodological issues might be responsible for this finding. Specifically, following work on English by Gualmini (2003), I suggested that the use of negation in the experimental scenario employed here might be infelicitous and this might have contributed to the number of non-targetlike responses on this task and, in particular, to the likelihood of L1 transfer on the part of the L2 subjects. Furthermore, the fact that the two tasks required different answers, that is that the stories employed in the tasks verified different readings, may also have caused differences in subjects' responses.

6.6.2 Comparison of truth value judgement task with act-out task

This section, which compares the truth value judgement task on *twee keer* with the act-out task on *twee keer*, concerns the L2 adults only. Recall that the truth

---

[40] The following modification to the experimental scenario used here would serve this purpose. Recall that in the *twee keer* task, the main character manipulates two (different) objects out of three. In the example scenario given in (14) the girl thus tickles a brown monkey and then a grey monkey. If in the final picture, the girl tickled the grey monkey for a second time (because he enjoyed it the first time so much, for example), the scrambled test sentence would be true on the specific reading (because there is a monkey which the girl tickled twice, namely the grey one) and false on the non-specific reading (because it is not the case that the girl tickled a(ny) monkey twice, rather she engaged in three monkey-tickling events). Carrying out this task will be left for future research.

[41] It is also possible that negation is conceptually harder, although this is only likely to cause a real problem for the younger subjects.

value judgement task was designed to ascertain whether subjects could correctly rule out the non-specific reading for the scrambled indefinite; the act-out task was used to see whether they demonstrated the same preferences as native adults, that is, whether they manipulated one object in response to a scrambled sentence and two objects in responses to a non-scrambled sentence. By comparing subjects' responses on these two tasks, it is possible to determine whether subjects who, by manipulating one and two objects, respectively, demonstrate knowledge of the specific meaning of the scrambled object and of the preferred interpretation of the non-scrambled object in the act-out task also rule out the non-specific meaning for the scrambled object. Additionally, it will be possible to ascertain whether those subjects whose 1-object responses in the act-out task were said to be uninformative stem from targetlike knowledge of the interpretive constraints on scrambling or from L1 transfer.

There were six subjects whose responses on the act-out tasks indicated that they made the targetlike distinction between scrambled and non-scrambled objects, that is, they consistently manipulated one object in the scrambling condition and two objects some or all of the time in the no scrambling condition. Five of these subjects were from the high proficiency group and the other was from the low proficiency group. On the truth value judgement task, this latter subject (A31C) consistently accepts the non-specific reading for the scrambled object, that is, there is a discrepancy between his responses on the two tasks. The five high proficiency subjects, however, make the targetlike distinction in the truth value judgement task as well. Four of them (A20C, A24C, A26C and A28C) always reject the non-specific reading for the scrambled object and always accept it for the non-scrambled object. The fifth one (A27C) rejects the non-specific reading for the scrambled object in three out of five items and always accepts it for the non-scrambled object. These subjects clearly have targetlike knowledge of the interpretive constraints on scrambling.

What about those subjects whose 1-object responses on the act-out task were uninformative? There were six subjects who consistently produced 1-object responses in both conditions in the act-out task. By themselves, these responses are uninformative as they could stem from targetlike knowledge or from transfer from the L1, where the preferred interpretation of sentences such as 'You should throw a ball twice' is also a 1-object response. By examining these subjects' responses on the truth value judgement task with *twee keer*, it will be possible to distinguish between these two options. On the truth value judgement task, two of the six subjects (A29C (low) and A23C (high)) incorrectly accept the non-specific reading for the scrambled object, suggesting that their 1-object responses in the act-out task were probably the result of L1 transfer. The other four subjects correctly reject the non-specific reading for the scrambled object some (A25C (low)) or all (A19C (mid), A21C (low) and

A30C (high)) of the time, which suggests that their 1-object responses in the act-out task were probably the result of targetlike knowledge of the interpretive constraints on scrambling.[42]

A comparison of the act-out task with the truth value judgement on *twee keer* thus indicates that ten of the 14 adult L2 subjects who took part in the act-out task have targetlike knowledge of the interpretive constraints on scrambling.

6.7 Summary

This chapter used three different tasks to track the development of scrambling in three different groups of learners, L1 children, L2 children and L2 adults. In L1 acquisition, the development of targetlike comprehension of scrambled indefinite objects was found to be a long and gradual process. Adultlike levels of targetlike responses were not attained until as late as 12 years old. A similar – though not completely identical – pattern was observed in responses to two different truth value judgement tasks, one testing scrambling across negation and one testing scrambling across the adverb *twee keer* 'twice'. The child L2 data collected using this latter task were suggestive of similar, age-related development: the younger children (at time of testing) were significantly different from the older children and natives, and unlike the high proficiency L2 adults, the high proficiency L2 children were significantly different from the native adults. Both groups showed an increase in targetlike responses with proficiency. A subset of the L2 adults were also tested on scrambling across *twee keer* using an act-out task. On this task, the high proficiency subjects demonstrated similar preferences as the native adults, indicating that they had complete targetlike knowledge of the interpretive difference between scrambled and non-scrambled indefinite object NPs. For the L2 children and adults, the truth value judgement *twee keer* data contrast sharply with those in the truth value negation data, where no clear patterns were found: there was no effect of proficiency or of age at time of testing. It was suggested that the combination of a potentially infelicitous use of negation in the experimental scenario and L1 transfer may be responsible for this. Further research is necessary to confirm whether this is the case.

The L1 results and the L2 results on *twee keer* are compatible with Krämer's (2000) Non-Integration Hypothesis, which states that limited discourse integration, which is necessary for the targetlike interpretation of scrambled indefinites, is the cause of the observed patterns. Although there is some evidence (e.g. Karmiloff-Smith 1979) to suggest that discourse integration in language comprehension is still developing at this age, further

[42] The one remaining subject, A18C, produces targetlike responses on the truth value judgement task and non-targetlike responses on the act-out task, where he consistently manipulates one object in the no scrambling condition and two objects in the scrambling condition. It is unclear how these data should be interpreted.

research is necessary to determine which aspects of discourse integration are causing children problems and under which circumstances, if any, discourse integration improves. Given that children of this age are clearly capable of successfully integrating discourse when completing other tasks (such as reading relatively complex stories such as *Harry Potter*, for example), it is not the case that they are always unable of integrating discourse. Other possible explanations for the observed late development in the targetlike comprehension of scrambled indefinites are explored in the following chapter.

The results of the *twee keer* task provide clear evidence that L2 children and adults are able to overcome the poverty-of-the-stimulus problem which the acquisition of the interpretive constraints on scrambled objects presents. They know that scrambled indefinites in Dutch cannot receive a non-specific interpretation despite this knowledge not being available in the TL input, via L1 transfer or TL instruction. On this task, the high proficiency L2 adults did not differ significantly from the native adults and although there was a significant difference between the high proficiency L2 children and the native adults, there were nevertheless several L2 children who provided targetlike responses.

In the following chapter, the results of the comprehension and production of scrambled objects will be compared. There, I also speculate as to how the interpretive constraints on scrambled indefinites might be acquired.

CHAPTER 7

PRODUCTION AND COMPREHENSION COMPARED

Introduction

The previous two chapters presented data on the production and comprehension of scrambled object NPs. The production experiment tested whether L1 children, L2 children and L2 adults were able to scramble definite and indefinite objects across negation in the appropriate contexts. The comprehension tasks determined whether these three learner groups assigned indefinite objects scrambled across negation and *twee keer* 'twice' a targetlike, specific interpretation. In this chapter, the production and comprehension data will be compared. Previous research on L1 acquisition suggests that there is a discrepancy between production and comprehension, that is, targetlike production of scrambled indefinite objects is in place before targetlike comprehension of scrambled indefinite objects. The production and comprehension results from the 12 individual L2 children who participated in both experiments confirm this pattern. After exploring various possible explanations for this finding, including methodological factors and limited discourse integration, I speculate as to how the interpretive properties of scrambled indefinite objects are acquired.

The chapter is organised as follows. Section 7.1 recapitulates the group findings for the L1 and L2 children from the previous two chapters. Subsequently, in §7.2, the individual child L2 results are presented. In §7.3.1, various methodological factors which might have contributed to the observed result are considered and in §7.3.2, other explanations are explored. Finally, in §7.4, an account of how all three groups might acquire the interpretive constraints on scrambled indefinites is proposed. A summary and conclusion follows in §7.5.

7.1 Recap: Child L1 and child L2 group results

The data presented in Chapter 5 demonstrated that there are 5-year-old L1 children who know the interpretive constraints on scrambled indefinites, that is, they scramble indefinite objects in a context where a specific interpretation is intended and they do not scramble them in a context where a non-specific interpretation is intended. These results contrast with those in Chapter 6, which showed that the comprehension of scrambled indefinite objects may not be targetlike (for all children) until as late as age 12. It thus appears that in the L1 acquisition of scrambled indefinite objects, there is a discrepancy between production and comprehension.

For L2 children, it was observed that the developmental stages in production subsume those found for L1 children. The difference between L2 and L1 children was in the initial child L2 stage, characterised by the order Negation-Verb-Object and analysed as the instantiation of L1 transfer. Targetlike production of scrambled indefinites was more or less in place by the mid proficiency level: L2 children in this group produced on average 71.7% scrambled objects, which compares with an average rate of 61.3% scrambled indefinite objects for the 5-year-old L1 children (cf. 91.7% in the high proficiency child L2 group). The child L2 comprehension data were suggestive of a similar age-related pattern as in the child L1 data, at least for scrambling across *twee keer*: the younger children generally (incorrectly) allowed the non-specific interpretation for the scrambled indefinite object, whereas the older children were less likely to do so, and unlike the high proficiency L2 adults, the high proficiency L2 children differed significantly from the native adults. Taken together, these results suggest that, as with the L1 children, in the child L2 acquisition of scrambling of indefinite object NPs, there is a discrepancy between production and comprehension.

The child L1 and most of the child L2 production and comprehension data discussed thus far have been from different children. There is, however, a subset of L2 children who participated in both tasks. The data from these children can be used to determine whether the production ~ comprehension discrepancy suggested above on the group level also holds for individuals. If this is the case, we might expect to find children who are targetlike on production but not on comprehension, but not the other way round.

7.2 Individual child L2 results on production and comprehension

In total, there were 12 L2 children who were tested on both experiments, that is, on the production task (on scrambling across negation) and on both truth value judgement tasks (on scrambling across negation and scrambling across *twee keer* 'twice'). The children were taken from one school, towards the end of the whole data collection procedure, when it became apparent that such a within-subject production ~ comprehension comparison could be important. Mid to high proficiency subjects (as determined by their teacher) were selected for participation. The rationale behind this was that low proficiency subjects were likely to provide non-targetlike responses on both tasks, which would be completely uninformative. The production task was completed first and the comprehension tasks in a second, separate session, within two weeks of the first. The comprehension task was carried out second to ensure that it did not prime subjects for the production task. I wanted to avoid the possibility that the test items in the comprehension task containing scrambled objects would somehow help subjects to provide targetlike scrambled responses in the production task.

The results for all three tasks are given in Table 60. Subjects are assigned a secondary code for the purposes of this chapter, given in the first column. The second column gives, for the purposes of comparison, the codes which are used in the Appendices and the third column states the proficiency level. L indicates low proficiency, M designates mid proficiency and H high proficiency. Recall that subjects completed a picture description task as part of each experiment i.e. one for the production experiment and one for the comprehension experiment. The subjects reported on here therefore have two proficiency scores. Given that the proficiency scores are standardised scores based on the L2 group as a whole and that the majority of subjects were different for each experiment, these subjects' scores on the two proficiency tasks are not the same. As the data in the third column indicate, however, most (10/12) of these subjects were nevertheless assigned to the same proficiency level on the basis of each of the two scores. The two exceptions are subjects F and G: on the production and comprehension tasks, respectively, subject F scored –0.34 and –0.51 and subject G –0.32 and –0.66, placing them both in the mid proficiency level on the first task and the low proficiency level on the second. The remainder of Table 60 presents the results on the various tasks. Black shading indicates that a subject always provides targetlike responses (i.e. 100%), grey shading that targetlike responses are provided sometimes, and no shading represents consistently non-targetlike responses (i.e. 0%).

Table 60. L2 children: Production and comprehension compared (organised by age at time of testing)

	Subject Code	Prof. level	Age gp (at testing)	Production: Negation: Specific condition: % objects scrambled	Production: Negation: Non-specific condition: % objects scrambled	Comprehension: Negation: Scrambled condition: % true answers accepted	Comprehension: *Twee keer*: Scrambled condition: % false answers rejected	Comprehension: *Twee keer*: No scrambling condition: % true answers accepted
A	C18P/ C17C	L	8	100% (5/5)	100% (2/2)	100% (5/5)	0% (0/5)	100% (5/5)
B	C21P/ C20C	H	8	40% (2/5)	0% (0/2)	0% (0/5)	0% (0/5)	100% (5/5)
C	C24P/ C25C	H	8	100% (6/6)	0% (0/4)	0% (0/5)	0% (0/5)	100% (5/5)
D	C17P/ C15C	M	8	100% (3/3)	50% (2/4)	100% (5/5)	0% (0/5)	100% (5/5)
E	C23P/ C22C	H	9	100% (4/4)	0% (0/1)	0% (0/5)	100% (5/5)	100% (5/5)
F	C19P/ C18C	L/ M	9	80% (4/5)	100% (4/4)	0% (0/5)	100% (5/5)	0% (0/5)
G	C14P/ C11C	L/ M	10	50% (3/6)	0% (0/4)	100% (5/5)	0% (0/5)	100% (5/5)
H	C25P/ C26C	H	10	75% (3/4)	0% (0/5)	40% (2/5)	100% (5/5)	0% (0/5)
I	C22P/ C21C	H	10	100% (6/6)	0% (0/2)	100% (5/5)	100% (5/5)	0% (0/5)
J	C16P/ C14C	H	11	100% (6/6)	100% (5/5)	100% (5/5)	0% (0/5)	100% (5/5)
K	C15P/ C12C	H	11	100% (1/1)	0	0% (0/5)	0% (0/5)	100% (5/5)
L	C20P/ C19C	M	11	100% (5/5)	0% (0/2)	100% (5/5)	0% (0/5)	100% (5/5)

Let us first consider the production data for these 12 subjects, given in the fifth column for the specific condition and the sixth column for the non-specific condition. Subject K produces too few data in this task to allow a comparison with the comprehension tasks and hence he is excluded from further discussion. All of the remaining subjects produce scrambled objects some or all of the time and most of them make a (relative) distinction between the two conditions, scrambling in the specific condition but not in the non-specific condition. There are however, three subjects (A, F and J) who scramble in the non-specific condition. As noted in §5.4.3.4, scrambling in this condition is unexpected rather than ungrammatical. Although the task was not designed to elicit a specific reading for the indefinite object, there is nothing to prevent the subject from doing so. We can therefore not exclude the possibility that the specific meaning was intended.[1]

Comparing the production data with the comprehension data on negation (in the seventh column), the following patterns are observed:

(a) Always targetlike production but non-targetlike comprehension: subjects C and E

(b) Sometimes targetlike production but non-targetlike comprehension: subjects B and F

(c) Always targetlike production and always targetlike comprehension: subjects A, D, I, J and L

(d) Sometimes targetlike production but always targetlike comprehension: subject G

(e) Sometimes targetlike production and sometimes targetlike comprehension: subject H

At first sight, it appears that every possible response pattern across the two tasks occurs. There is, however, one clear generalisation which can be made over these data, namely that subjects who (always or sometimes) have targetlike comprehension, (always or sometimes) have targetlike production. The opposite does not hold, however: subjects who (always or sometimes) have targetlike production do not (always or sometimes) have targetlike comprehension.

It was already observed in §6.6.1.1 that subjects' responses on the two truth value judgement tasks do not necessarily tally with each other and that targetlike responses on one task do not necessarily entail targetlike responses on the other. Nevertheless, until the potentially problematic use of negation in the negation truth value experiment has been addressed, the assumption remains that scrambling across negation and scrambling across *twee keer* essentially involve the same process. Bearing this in mind, the production data (on negation) will now be compared with the comprehension data from the scrambled condition in the *twee keer* task (column 8). The following patterns are observed:

---

[1] Note also that it is not the case that those subjects who unexpectedly scramble in the non-specific condition are the same ones who incorrectly reject the non-specific interpretation of the non-scrambled order in the *twee keer* task (the rightmost column in Table 60). This is what would be expected if, for example, these unexpected/non-targetlike patterns were the result of an underlying grammar which disallowed the <non-scrambled, non-specific> pairing altogether.

(a) Always targetlike production but non-targetlike comprehension: subjects A, C, D, J and L

(b) Sometimes targetlike production but non-targetlike comprehension: subjects B and G

(c) Always targetlike production and always targetlike comprehension: subjects E and I

(d) Sometimes targetlike production but always targetlike comprehension: subjects F and H

Again, although a variety of patterns are observed in the data, the same generalisation holds here as it did for the negation comprehension data, namely that subjects who (always or sometimes) have targetlike comprehension, (always or sometimes) have targetlike production, but those who (always or sometimes) have targetlike production do not (always or sometimes) have targetlike comprehension. The existence of pattern (d) might suggest that there are subjects who are better on comprehension than on production, which would seem to contradict this generalisation. The two subjects with this pattern, F and H, produce scrambled objects at a rate of 80% (4/5) and 75% (3/4), respectively, however, which contrasts only slightly with their 100% targetlike rejection rate of the non-specific reading for the indefinite object scrambled across *twee keer*.

In this subset of L2 children, there are slightly more subjects who are targetlike on the negation task than on the *twee keer* task: six out of the 12 subjects consistently accept the specific reading for the scrambled indefinite on the negation task and a seventh does so sometimes (a 'mixed' response pattern), and on the *twee keer* task, just four subjects correctly reject the non-specific reading for the scrambled indefinite. This contrasts with the general finding in Chapter 6 that the L2 children (and adults) were, on the whole, worse on the negation task than on the *twee keer* task. This might suggest that hearing the production task first somehow primed these children for the comprehension task. In particular, it could be the case that the scenarios used in the specific condition of the production task, where one object was left unmanipulated, might have made the subjects more sensitive to the presence of a final unmanipulated object in the scenarios used in the truth value judgement task on negation. It is not clear how likely this is given that there was generally a one to two week gap between the two experiments. Note, however, that even if the production task did have such a priming effect, the generalisation made above regarding the relationship between production and comprehension still holds. Furthermore, there is no way in which the children could have derived the targetlike form-meaning pairing from the experimental scenario in the production task and hence, we can be sure that they do have the relevant targetlike knowledge.

Eight L2 adults also participated in both experiments. Three of these (A32C/A20P, A33C/A21P and A36C/A22P) were ten years old at the time of testing and they were tested within the space of two weeks as described above for the L2 children.[2] For the remaining five (A20C/A18P, A22C/A11P. A25C/A14P. A26C/A8P and A28C/A17P), the gap between the two experiments was approximately 18 months, with the production experiment being carried out first. Unfortunately, these adult data are largely uninformative because there are too few tokens for half of the subjects and because three of the remaining four fall within the age range in which limited discourse integration may play a role. For the sake of completeness, however, the details are provided here. In the specific indefinite condition, four of the eight subjects produced either one indefinite object only or none at all, making a comparison of the two tasks impossible. The remaining subject (A28C/A17P) had targetlike production (2/2 scrambled indefinites in the specific condition) and targetlike comprehension on the *twee keer* task, but not on the negation task. Of the three younger L2 adults, subject A36C/A22P had targetlike production (5/5 scrambled indefinites) and targetlike comprehension on the negation task, but not on the *twee keer* task, subject A33C/A21P had targetlike production (5/5 scrambled indefinites) but non-targetlike comprehension on both tasks and subject A32C/A20P had non-targetlike production (0/5 scrambled indefinites) and targetlike comprehension on both tasks. This latter subject is the only one who does not fall within the generalisation made above, that is, he is the only subject who has targetlike comprehension but non-targetlike production. More data on both tasks from the same L2 adult subjects, and in particular, from older L2 adults, are needed in order to determine the extent to which this subject is an exception or whether L2 adults follow a different pattern.

7.3 Discussion

The individual child L2 results examined here are consistent with the observation made above for the child L1 and child L2 group data that there is a discrepancy between the production and comprehension of scrambled indefinite objects, with targetlike production being in place before targetlike comprehension. Note, however, that in these data, there are no subjects who *fail* to produce scrambled objects. The claim that there is a production ~

[2] Recall that the definition of an L2 adult adopted in this thesis is 'a non-native acquirer whose initial exposure to the target language is at an age of eight years or older' (cf. §1.1). This does not exclude the possibility that there will be L2 adults who are relatively young at the time of testing. How these 'younger' L2 adults may differ from the 'older' L2 adults is discussed in §6.3.3.1 and §6.4.3.1, where the adult L2 comprehension data are split into two groups.

comprehension discrepancy would be further strengthened if it could be shown that all subjects who fail to produce scrambled objects also fail to interpret them correctly.

The observation that production in some sense might precede comprehension is surprising given that the reverse pattern is generally considered to be the norm (see e.g. Smolensky 1996 for relevant discussion). This raises the question of whether the observed discrepancy between production and comprehension is real or whether the delay in targetlike comprehension can be attributed to some other factor, for example, some aspect of the experimental methodology. In the remainder of this chapter, I first examine some such methodological factors and assuming on the basis of the data currently available that there is indeed something requiring further explanation, I subsequently evaluate possible accounts of the observed production ~ comprehension discrepancy. In the final section, I go on to speculate about how the interpretive constraints on scrambled indefinite objects might be acquired.

7.3.1 Methodological factors

This section considers whether it is possible that the difference between the production and comprehension data is a methodological artefact. More specifically, it examines whether the apparent delay in targetlike comprehension of scrambled indefinite objects could be accounted for by the experimental method and design which were employed. There are three areas of concern: the salience of the different objects in the experimental scenarios, the mode of presentation and the potentially infelicitous use of negation. The latter issue was examined in detail in §6.6.1.3 and will not be discussed any further here. The other two issues are addressed in turn below.

7.3.1.1 Salience of different objects

In a recent study, Lidz and Musolino (submitted) ask whether the fact that in their earlier experiments, the objects most relevant to the non-specific interpretation of sentences such as (1) might be considered more salient in the experimental scenario than the ones which are relevant for the specific reading, and whether this may have led the children to reject the specific reading in favour of the non-specific reading.

(1) Cookie Monster didn't eat two cookies

In particular, for the sentence in (1), a scenario was used where out of four available cookies, Cookie Monster ate two and left two. The fact that the two objects which are manipulated (i.e. eaten) are arguably more salient than those which are not could have led children to reject (1) (on the grounds that Cookie Monster *did* eat two cookies – the non-specific reading), even though this sentence is actually true in this context (because there are two cookies which

Cookie Monster did not eat – the specific reading). This is arguably also a problem in the type of scenario used in Krämer's negation experiment, which was replicated here. The two fish which are caught are arguably more salient than the one which is not caught. Lidz and Musolino argue that if the saliency of the objects affected children's responses in this way, we would expect that children would also reject sentences with numerically quantified subjects, as in (2), when presented in a similar context.

(2) Two butterflies didn't go into the city

In a context where out of four butterflies, two go to the city and two do not, children would be expected to reject (2) (because two butterflies *did* go to the city – the non-specific reading), even though it is true in this context (because there are two butterflies who did not go to the city – the non-specific reading). This is not what children do, however. Lidz and Musolino find that children have no problem accepting sentences such as (2) in the given context and hence they conclude that the salience of the objects is unlikely to be responsible for the children's responses to sentences with numerically quantified objects, such as (1). It is not clear, however, whether such a direct analogy between subjects and objects is possible, as these differ in several non-trivial ways (see de Hoop and Krämer to appear for relevant discussion). Furthermore, it is possible that this conclusion cannot straightforwardly be extended to the experimental scenario used here as it differs from both of Lidz and Musolino's experiments in that the number of objects which is manipulated does not equal the number of objects left unmanipulated. Here, there are two objects which are manipulated and one which is not. This difference might plausibly increase the saliency of the unmanipulated objects and thus of the non-specific reading.

In a follow-up experiment to the negation task, Krämer (2000) altered the set-up in the original task such that the focus of the story was on the set of objects to be manipulated. Thus, the first picture excluded the protagonist and focussed on the objects only. For example, in the story about the boy with the fish, the first picture contained a picture of the fish in the pond only (cf. example (3) in Chapter 6). Also, the unmanipulated object in the final picture was explicitly named: 'This fish is still swimming in the water and the other fish are gone.' Highlighting the set of objects in this way and in particular, the one object which had been left unmanipulated failed to have a significant effect on the number of targetlike responses produced by the children, however. On average, children accepted the specific interpretation 26.6% (38/143) of the time in the follow-up experiment compared with the 16.0% (25/156) in the main experiment. There was, nevertheless, a slight improvement for the 7-year-olds (6;10-7;10), who scored 44.4% (24/54) on the follow-up experiment compared with 10.6% (7/66) on the original. The 4-year-olds (4-5;6) were less targetlike on the follow-up experiment (0% (0/42) vs. 16.7% (7/42)) and for the 5/6-year-olds (5;6-6;10), there was hardly any change (29.8% (14/47) vs. 22.9% (11/48)). Thus, Krämer found that making the objects and in particular the unmanipulated object more salient did not significantly improve children's responses. However, even in the modified story, considerable attention was still paid to the two fish that were manipulated – they were the only things depicted in two of the four pictures. The possibility that this may have had an effect, as suggested above, cannot be excluded.

Krämer's follow-up experiment and Lidz and Musolino's experiment suggest that increasing the saliency of the unmanipulated object(s) does not, on the whole, lead children to provide more targetlike responses, although as just noted above, more data are needed. Both of these studies concern the specific interpretation of indefinite objects in negative sentences only, however. The data collected here also concern scrambling across the frequency adverb *twee keer*. It is plausible that the relative salience of the objects in the experimental scenario employed in the *twee keer* task may have had an unfavourable effect on the results. Recall that on this task, subjects were presented with stories where the main character manipulates two out of a group of three objects. For example, a girl tickles two different monkeys, a black one and a grey one, from a set of three. The scenario thus depicts the non-specific interpretation of the indefinite object. There is no object visible to the subject which is manipulated on two separate occasions, that is, the specific interpretation of the indefinite object is not given in the experimental scenario (cf. the negation experiment). In order to provide a targetlike response, subjects are therefore required to consider an alternative scenario which is not given to them. In this sense, the objects relevant to the non-specific interpretation might be considered more salient. In order to test whether this difference has an effect on subjects' responses, a follow-up experiment where the experimental scenario is consistent with the specific reading (see footnote 40 in Chapter 6 for a suggestion as to what such an experiment might look like). This will be left for future research.

7.3.1.2 Mode of presentation

The second methodological factor which may be responsible for the discrepancy between the production and comprehension results is the way in which the stories in the comprehension tasks were presented. Most of the previous experiments on the acquisition of scope (cf. §3.3.1.3) use a truth value judgement task where the stories are acted out using toys in front of the child, rather than using pictures as in the experiment used here and in Krämer (2000). It is possible that the use of such static pictures might have an effect on children's behaviour. Miller and Schmitt (2004:324) note this and in order to control for this possible effect, they acted out their experiment on indefinite objects embedded under negation (see below for details) with toys instead of using pictures. They find little difference between the two sets of results, however, and if anything, children are slightly worse when the stories are acted out than when they are presented with pictures. With respect to the present experiment, this suggests that the mode of presentation is unlikely to be responsible for the delay in comprehension compared with production.

7.3.1.3 Summary

There are two methodological concerns which require attention before we can be fully satisfied that the production ~ comprehension discrepancy observed for the acquisition of the interpretive constraints on scrambling really holds. These are the potentially infelicitous use of negation and (possibly) the salience of the objects relevant to the different readings of the indefinite NP.

7.3.2 Possible explanations for production ~ comprehension discrepancy

Putting the methodological concerns outlined in the previous section to the side for the moment, let us assume that the pattern observed in the group and individual data here is real and that there is indeed something which requires an explanation. In this section, I will consider two possible ways in which this pattern might be accounted for, namely limited discourse integration, as put forward in Krämer's (2000) Non-Integration Hypothesis and further developed in Miller and Schmitt (2004), and the less controlled nature of comprehension as discussed in Hurewitz, Brown-Schmidt, Thorpe, Gleitman and Trueswell (2000).

7.3.2.1 Limited discourse integration

In Chapter 6, it was argued that the child L1 and child L2 data were largely compatible with Krämer's Non-Integration Hypothesis, that is, children's non-targetlike interpretation of the scrambled indefinites until a relatively late age could plausibly result from limited discourse integration. However, given that for (some) children, targetlike production is in place at an earlier age, limited discourse integration cannot mean that children lack targetlike knowledge of the interpretive constraints on scrambling full stop. Rather, it seems that children have this knowledge underlyingly, but in comprehension, they sometimes fail to use it and when they do, they appear to resort to some sort of default option, where the indefinites are interpreted as predicatives, which is their basic type (de Hoop 1992; van der Does and de Hoop 1998; Van Geenhoven 1998). Krämer (2000:128) suggests that comprehension may lag behind production because 'the hearer's task of reconstructing the speaker's intentions is more complicated than the speaker's task of putting her intentions into words'. This idea is developed further in a recent paper by de Hoop and Krämer (to appear), which will be discussed in some detail below (§7.4.2).

Building on Krämer's (2000) work, Miller and Schmitt (2004) make a very explicit suggestion as to how limited discourse integration is responsible for children's failure to allow the specific interpretation of indefinite objects (in negative sentences). Adopting an analysis of indefinites put forward by Geurts (2002), they propose that children's failure to interpret specific indefinites in a targetlike fashion results from their inability to background the relevant set from which the indefinite (on this analysis, a covert partitive) object is taken. Miller and Schmitt use a comparable scenario to the one used here, where one object from a group of objects is left unmanipulated. They modify this scenario, however, such that the link between the unmanipulated object and the set from which it was taken is made clearer by using objects which form a more natural group and which are associated with a physical entity connecting them together, such as the candles on a birthday cake, the drawers in a chest etc. Additionally, in their experimental scenario, the unmanipulated object was also highlighted in a separate picture. They find that the L1 English 4-year-olds they tested correctly accepted the specific reading for sentences such as (3) in 92.3% (48/52) of cases.

(3) Timothy didn't blow out a candle

Miller and Schmitt claim that, in accordance with Krämer's Non-Integration Hypothesis, this significant proportion of targetlike responses results from their having provided children with explicit discourse referents, which made it easier for them to integrate discourse, and hence their responses were more targetlike than in experiments where this link is not provided. If limited discourse integration contributes to the observed delay in the child (L1/L2) Dutch data, it is expected that by altering the experimental scenarios accordingly, the number of targetlike responses produced by L1 Dutch children should also increase.[3]

7.3.2.2 Acquisition of the relevant cues and cue integration

In a series of production and comprehension experiments, Hurewitz *et al* (2000) investigate the L1 acquisition of restrictive modification. They show that while children are garden-pathed on sentences such as (4), interpreting the PP, 'on the napkin', as a destination rather than a modifier, the same children *are* able to produce PPs as modifiers, even in exactly the same context.

(4) Put the frog on the napkin into the box

Hurewitz *et al* suggest that this production ~ comprehension discrepancy results from problems children's experience with parsing. For them, parsing relies on the successful integration of a number of cues from different sources, such as visual context and the lexical preference of verbs; some of these cues are more reliable than others. It could be that children have acquired the more

---

[3] It should be borne in mind, however, that unlike the English sentence in (3), sentences with scrambled indefinites in Dutch are not ambiguous.

reliable cues, namely that the verb 'put' usually goes with a destination phrase, but not others, and because they are unable to maintain alternative analyses for an extended period of time, they are unable to retract from their first guess. In other words, children's delay in comprehension stems could stem from a delay in the acquisition of the relevant cues and/or the successful integration of those cues.

Hurewitz *et al*'s proposal raises the interesting issue of which cues Dutch provides, besides structural position, to indicate that the scrambled indefinite object should be interpreted specifically. The native adults in the production task tended to realise scrambled indefinite objects as the stressed indefinite/cardinal *één* rather than the unstressed indefinite determiner *een*. If this is in any way representative of normal spontaneous speech, the form of the indefinite might also be considered a cue (albeit not an unambiguous one) for the targetlike interpretation of scrambled indefinites.[4] The comments made by the L2 subjects, especially the adults, suggest that the form of the indefinite is definitely a cue for them (see §5.4.3.3 for relevant discussion). The discussion in the preceding section suggests that surrounding discourse is also an important cue, especially for the L1 and L2 children. It could be that in addition to its scrambled position, native adults also make use of the form of the indefinite (*één* vs. *een*) and the surrounding discourse, but unlike L1 and L2 children, when these cues are not available, they can still access the target specific interpretation. It could also be that children are not able to 'fill in' the missing cues in the same way as native adults can. It is also possible that children first latch onto the form of the indefinite as an indicator of a specific interpretation and when this is not available, parsing fails. A simple way to test this would be to replicate the comprehension experiments used in Chapter 6 with the indefinite object realised as *één* rather than *een* in both scrambled and non-scrambled sentences. This will be the subject of future research.

It is unclear to what extent Hurewitz *et al*'s proposal could account for the scrambling data. More research on the cues involved in parsing scrambling by native adults and language learners is required. What is clear, nevertheless, is that to say that children have not acquired the most reliable cue, the position of the NP, cannot be right given that they use this in the target contexts in production.

7.3.2.3 Summary

This section explored two possible explanations for the observed difference between the production and comprehension of scrambled indefinite objects by L1 and L2 children. One possibility is that non-targetlike comprehension is due to limited discourse integration, which, on Krämer's (2000) approach, could in turn be the result of processing limitations, or on Miller and Schmitt's (2005)

---

[4] This assumes that *één* is in fact a stressed indefinite determiner rather than a cardinal. See below and §2.1.2 for relevant discussion.

account, the inability to identify the set to which the object in question belongs. Another option is that children use different cues to identify the target specific reading of the indefinite object and/or they are unable to integrate all the relevant cues. In order to explore these issues in more detail, more data are needed. In particular, more research is needed on children's processing abilities and how exactly these relate to the comprehension of scrambling. Crucial to future research, I believe, will be the collection of production and comprehension from the same L2 adults. As noted above, although a handful of adults were tested on both tasks, the data obtained were too few to be able to say anything meaningful. If the delay in comprehension observed here is the result of limited discourse integration, the prediction for L2 adults would be that a similar delay should only be observed in those adults who have limited discourse integration, that is, younger and/or low proficiency L2 adults. I return to this issue below.

7.4 How the interpretive constraints on scrambled indefinites might be acquired

The previous section reviewed two possible explanations for the observed delay in targetlike comprehension of scrambled indefinite objects in L1 children and English-speaking L2 children. The finding that at an earlier stage, children are able to produce scrambled and non-scrambled indefinite objects in the relevant contexts (*viz.* specific and non-specific, respectively) indicates that underlyingly, they must have targetlike knowledge of the interpretive constraints on scrambling. In this section, I speculate about *how* the targetlike, specific interpretation of scrambled indefinites might be acquired.

In Chapter 3, it was argued that for the English-speaking (child/adult) L2er, the acquisition of the interpretive constraints on scrambled (indefinite) objects constitutes a poverty-of-the-stimulus problem. Knowledge of the target specific interpretation of scrambled indefinites can neither be induced from the input nor can it be transferred from these L2ers' L1, English. Furthermore, it is not the subject of classroom instruction. Consequently, if English-speaking L2ers come to know these restrictions, their interlanguage grammars must somehow be constrained in the same way as native grammars. The question to be addressed here, then, is how do they do this? I speculate that they make use of a 'Blocking Principle', in the spirit of Williams (1997), amongst others.

7.4.1 The Blocking Principle

The general idea behind blocking, which appears under various guises in the literature (e.g. Elsewhere Condition (Kiparsky 1973); Unique Entry Principle (Pinker 1984); Uniqueness Principle (Pinker 1986; Wexler and Culicover 1980); Blocking (Marcus, Pinker, Ullman, Hollander, Rosen and Xu 1992)) is that different morphological or syntactic forms must have different meanings.[5] Williams (1997) notes that in this regard, the notion of specificity is a significant factor. It is important here to make a distinction between specificity in the sense that X is specified in more detail than Y and specificity in the sense that X refers to a particular (specific) entity. It is the former which is intended here. As a principle constraining the lexicon, blocking was originally defined as 'the non-occurrence of one form due to the simple existence of another' (Aronoff 1976:43). For example, in the derivation of nouns from adjectives, underived nouns, such as *glory*, are considered to block nominalisations derived by means of a general rule, as in *glorious* → **gloriosity*, as a consequence of the latter being the less specific (or the more general or elsewhere) form. Perhaps the most well-known application of blocking in L1 acquisition is in Pinker's (1984) work on the acquisition of the past tense marker, -*ed*, in English, where it is dubbed the Uniqueness Principle. Pinker notes that in the acquisition of past tense forms, children overgeneralise the productive inflectional rule (verb + *ed*) to irregular verbs, resulting in non-targetlike forms such as *breaked*. He suggests that once, on the basis of the input, children acquire the irregular form *broke*, this more specific form blocks the application of the general rule. In this way, children come to know that *breaked* is ungrammatical without having to make recourse to negative evidence.

Williams claims that the Blocking Principle is a principle of language acquisition (as well as a broad organising principle of grammar).[6] He notes that in a certain sense, it can be equated with Chomsky's (1991) economy principles or the principles of Optimality Theory (Prince and Smolensky 1997). (The similarities and differences – of which there are several – between the way in which the blocking is implemented here and how it applies in OT will be addressed below, in light of a recent OT account for the acquisition of scrambling put forward by de Hoop and Krämer (to appear).) Although the classic examples of blocking concern the lexicon, Williams shows that blocking relationships are also found in other domains.[7] Let us consider an example from Turkish. In that language, case marks specificity (Enç 1991), as illustrated in (5).

---

[5] I do not wish to claim that these various proposals are completely identical, rather that they capture the same general idea.
[6] See Van Kampen (2004) for a recent application of the Blocking Principle in L1 acquisition.
[7] See Poser (1992) for a similar observation, albeit followed by a different conclusion.

(5) a. Iki kız-ı tanıyordum (partitive)
two girls-ACC I-knew
'I knew two (particular) girls.'

b. Iki kız tanıyordum (existential)
two girls I-knew
'I knew two girls'

The object in (5)-a, marked with the case marker *–i* receives a partitive (specificity) interpretation, whereas the object in (5)-b, where case is not marked, is interpreted existentially (non-specifically). The alternation in (5) exemplifies two of the characteristics of a blocking relationship between two forms, namely (i) complementarity and (ii) a special condition on one of the two forms. Thus, form A occurs where form B does not and the two are distinguished from each other in some way; in this example, the special condition on one of the forms is the application of case-marking, but it could just as easily be the application of a morphological rule or syntactic movement, etc. Blocking associates two different dimensions and although the choice of which dimensions associate with each other is arbitrary, the way in which they are associated is not: Williams notes that it always seems to be the case that the more specified ends of the two dimensions are associated with each other. The two dimensions which are associated with each other in (5) are thus case-marking (morphosyntax) and meaning (semantics) and the more complicated form is thus associated with what is assumed to be the more complicated interpretation (see below for further discussion). Blocking applies in the following way: the case-marked form is not associated with the non-specific interpretation because for this meaning, a less complex form (or 'cheaper') form exists, namely the non-case-marked form.

I believe that this application of the Blocking Principle can be extended to the acquisition of the interpretive properties of scrambled indefinite objects in Dutch, both by L1 children and by English-speaking (child/adult) L2ers.[8] The learnability problem faced by these three groups of learners is the same. At the initial state, learners in all three groups do not have any knowledge of scrambling and the L2ers cannot transfer this from their L1, because English does not have scrambling. There is nothing in the target language input to tell the learners that the scrambled form is not simply a rewrite of the non-scrambled form.[9] In other words, although the L2 input will provide them with situations where the scrambled form is used in a situation where a specific

[8] This section is a development of the ideas outlined in Unsworth (2005).
[9] In Chapter 3, I demonstrated that the two circumstances where the input could potentially provide the learner with the relevant information were likely to be very infrequent and, more importantly, not always unambiguous (cf. §3.4).

interpretation of the indefinite object is required, it will not provide them with *<scrambled, non-specific> pairings.[10, 11]

I would like to propose that in order to overcome the poverty-of-the-stimulus, learners make use of the Blocking Principle in the following way. The association between the scrambled form and the non-specific interpretation is ruled out because this <form, meaning> pair is blocked by the existence of the <non-scrambled, non-specific> pair. To put it slightly differently, once the learner knows that scrambling is an option in Dutch, that is, once the interlanguage grammar allows direct objects to be moved from their base position to a scrambled position, the non-specific interpretation will not be associated with the scrambled indefinite because for that interpretation of the indefinite, an alternative, more 'basic' or less 'costly' form, the non-scrambled form, exists. The non-scrambled form can be considered 'cheaper' or 'more basic' because it does not involve syntactic movement.[12] This follows from the idea that a derivation which does not involve movement is less costly than a derivation which does (i.e. economy construed globally in the sense of Adger 1994; Reinhart 1995). This less 'costly' or 'complex' form associates with the 'less complex' or basic type for the indefinite object, namely the predicative type (de Hoop 1992; van der Does and de Hoop 1998; Van Geenhoven 1998). In other words, the more specified form (the scrambled order) associates with the more specified meaning (the free variable or <<*e*,*t*> *t*> type), as is one of the characteristic properties of a blocking relationship. The free variable and <<*e*,*t*> *t*> types can be considered more specified in the sense that they result from the application of some special operation (such as type-shifting).[13]

---

[10] Learners would be forced to rely on (indirect) negative evidence. According to Philip (2003), this is exactly what learners do; the data presented in Chapter 6 were found to be incompatible with this account, however.

[11] Note, however, that there is one important difference between L1 children, on the one hand, and L2 children and adults, on the other, and that is the following: even if English-speaking L2 children and adults fail to interpret scrambled objects in a targetlike fashion, they should still have access to the specific interpretation of the indefinite object in their L1 (Miller and Schmitt 2004), whereas when young L1 Dutch children fail to assign the target specific interpretation to the scrambled object, this could be because they have not acquired this interpretation yet.

[12] The only way I can envisage this proposal working on a base-generation approach to scrambling is by making recourse to the (in)frequency of scrambled indefinite objects. It is not clear however how language learners should come to know such frequency facts or how they should be implemented in the grammar.

[13] The reader may recall that the non-scrambled form is also consistent with a specific interpretation of the indefinite object. This may seem at odds with the blocking proposal put forward here because a blocking relationship implies a 1-to-1 relationship between form and meaning. Recall, however, that the observation that the non-scrambled form is also consistent with the specific reading of the indefinite object is the result of entailment relations, not because this reading is genuinely available (see §2.1.2). The claim that there is no strict 1-to-1 relationship between form and meaning with respect to scrambling stems from the availability of intermediate as well as widest-scope readings for scrambled indefinite objects (§2.3.3.2). I believe that the application of the Blocking Principle put forward here is not incompatible with

One of the immediate questions that such a proposal raises is why blocking should apply in certain cases but not in others. For example, why is there no blocking relationship between scrambled and non-scrambled definite objects? Also, languages have numerous ways of expressing the same meaning with different forms – should we not expect a blocking relationship there, too? And is it not the case that blocking should mean that optional movement should not exist at all? I believe that the crucial factor which distinguishes these examples from the scrambling of indefinite objects (and presumably other properties of language which demonstrate a blocking relationship at the syntax-semantics interface) is the lack of a difference in (truth-conditional) meaning between alternate forms. Let us consider each example in turn.

As outlined in §2.1.1, the scrambling of definite object NPs is generally optional, as illustrated in (11) and (12), and it is not associated with a difference in meaning.[14] The sentences containing the scrambled definites, (11)-b and (12)-b, have the same truth conditions as their non-scrambled counterparts in (11)-a and (12)-a. In a particular context, there may be a preference for the scrambled over the non-scrambled order, for example, if the object is strongly anaphoric, but the two orders do not differ, semantically speaking.

(6) a. John heeft gisteren de boom geplant
John has yesterday the tree planted

b. John heeft de boom gisteren geplant
John has the tree yesterday planted
'John planted the tree yesterday.'

(7) a. Gert Jan heeft twee keer de bloem besproeid
Gert Jan has two times the flower watered

b. Gert Jan heeft de bloem twee keer besproeid
Gert Jan has the flower two times watered
'Gert Jan watered the flower twice.'

When negation is involved, this preference to scramble anaphoric NP objects interacts with the scope-taking properties of the negator, *niet*. Sentential negation is therefore expressed with the scrambled order, in (8)-a, whereas when the definite object is not scrambled, as in (8)-b, the interpretation is one of contrastive negation. There is a clear difference between the two, but once again, they do not differ truth-conditionally.

---

this fact as its purpose is to rule out the non-specific reading for the scrambled object and this is equally necessary when that object takes intermediate scope.

[14] See de Hoop, Haverkort and van den Noort (2004) for discussion of the Blocking Principle applied to scrambling and definite NP objects.

(8) a. Bob heeft het onkruid niet weggegooid
Bob has the weeds not away-thrown
'Bob didn't throw away the weeds.'
b. Bob heeft niet het onkruid weggegooid
Bob has not the weeds away-thrown
'Bob didn't throw away the weeds.'

The same (of course) holds for the often numerous ways which a language has to express the same meaning. For example, although the sentences in (9)-a and (9)-b differ from (9)-c in that they involve clefting and topicalisation, they have the same truth conditions as the simple declarative in (9)-c.

(9) a. It's the combination of orange and chocolate that I can't stand.
b. Orange and chocolate, that's the combination I can't stand.
c. I can't stand the combination of orange and chocolate.

At first blush, the existence of all three forms would appear to go against the notion of blocking outlined above because it might be expected that (9)-c, as the 'cheaper' form, should block (9)-a and (9)-b, but it does not. Recall that as it is implemented here, blocking serves to ensure that, in addition to the specific interpretation, the scrambled indefinite does not receive a non-specific interpretation, because for this interpretation, a cheaper (non-scrambled) form is available. For the analogy with the scrambling case to go through, (9)-a and (9)-b would have to be interpreted *differently* from (9)-c and as a consequence of blocking, the meaning associated with (9)-c would not be available. This is not the case, however, because although they clearly have different topic-focus structures, these three sentences do not differ truth-conditionally. That the difference appears to be one of information structure rather than semantics is comparable to the instances of definite object scrambling discussed above in (11), (12) and (8). It seems, therefore, that blocking does not apply in such cases. Rather, for blocking to occur at the syntax-semantics interface, it appears that truth-conditionally different meanings must be available as well as a difference in syntactic form.

The difference between the scrambling behaviour of definite objects and indefinite objects suggests that it might be the nature of the element which is involved in the movement operation which is crucial in this regard.[15] More in particular, this element should be ambiguous in some way, such that a difference in (truth-conditional) meaning is enabled. Indefinites are ambiguous between a specific and a non-specific reading, whereas this is clearly not the case for definites. Another example where blocking could apply, this time involving pronouns, is when PP-extraposition takes place in a sentence containing a pronoun, as in (10) (cf. Guéron 1980).

---

[15] As the Turkish example in (5) illustrates, a blocking relationship does not have to involve a movement relation. Case-marking could also be involved, for example.

(10) a. A picture of Jocelyn$_j$ was given to her$_{j,k}$ (coreferential/deictic)
b. A picture t$_i$ was given to her$_{*j,k}$ [of Jocelyn$_j$]$_i$ (deictic)

The pronoun in (10)-a is ambiguous between a coreferential and a deictic interpretation, that is, it can refer to either *Jocelyn* or to some other female person in the context. When the PP containing the proper name is extraposed, however, as in (10)-b, only the deictic reading is available, even though, as in (10)-a, there is nothing structural (such as c-command) which would prevent a coreferential reading. The coreferential reading could therefore be said to be blocked by the availability of a 'cheaper' form with the same meaning.[16] The prediction for acquisition is therefore that once L1 children or L2ers whose L1 does not instantiate the same pattern link these two forms by means of syntactic movement, that is, once they know that extraposition is an option in English, they should not allow the coreferential interpretation of the pronoun. In other words, all things being equal, targetlike production and targetlike comprehension should coincide.

Optional movement operations such as the one illustrated in (11) also follow the pattern observed in (11) through (9) above, namely that when there is no element which could lead to different meanings (such as an indefinite), blocking does not seem to apply.

(11) a. What did Eileen not do?
b. What didn't$_i$ Eileen t$_i$ do?

Williams (1997:579) notes that given that the various properties of language associated with each other by means of a blocking relationship are language-particular and largely idiosyncratic, the 'language acquisition mechanisms must be sufficiently sharp to extract these details from the primary linguistic data'. The Blocking Principle can therefore be viewed as a vital part of the language acquisition process, a language acquisition principle which exploits linguistic primitives, such as syntactic movement and semantic types, and which ensures that language learners are able to acquire the target language constraints in question.

Having addressed some of the issues concerning the domain of application of the Blocking Principle, we now return to the acquisition data. At what stage is blocking supposed to apply in the acquisition of scrambled indefinite objects? I propose that the acquisition of the interpretive constraints on scrambled indefinites is characterised by the following three stages:

[16] To say that the example in (10) involves a blocking relationship means that it must be assumed that the coreferential interpretation is less 'specified' or 'complex' than the deictic one.

(12) **Stages in the acquisition of the interpretive constraints on scrambled indefinite objects**

Stage	Underlying grammar …	How this manifests itself …
A	No knowledge of scrambling	Non-targetlike production and non-targetlike comprehension
B	Targetlike knowledge of scrambling coupled with limited discourse integration	Targetlike production and non-targetlike comprehension
C	Targetlike knowledge of scrambling and discourse integration	Targetlike production and targetlike comprehension

All three groups of learners, L1 children, (English-speaking) L2 children and L2 adults should go through stage A. At this stage, L2 children and adults will have knowledge of the specific interpretation of indefinite objects, but they do not know that in Dutch, this is expressed by means of scrambling. L1 children do not know this either and they may also not know that indefinite objects can be interpreted specifically. On the proposal put forward here, it is the Blocking Principle which allows learners to progress beyond stage A. Note that the claim that learners make use of the Blocking Principle does not necessarily mean that once they start scrambling indefinite objects, they should always scramble; the acquisition of the scrambled form may take time (cf. §5.4.3.2). What the Blocking Principle does predict, however, is that when learners scramble, they do so in the targetlike contexts only, that is, they should not scramble in a context where the indefinite object can only have a non-specific reading. More concretely, in describing a situation such as the one used in the negation experiment, where a boy fails to catch one fish from a group of three, it is expected that learners might sometimes scramble, but they will not necessarily do this consistently. However, in describing a situation such as the one used in the *twee keer* experiment, where a girl tickles two different monkeys, learners should not use the non-targetlike, scrambled form.[17]

Let us now consider stage B. Putting any methodological issues to one side for the moment, and on the assumption that processing problems in the early stages of L2 acquisition will mean that limited discourse integration will be a factor in adult L2 acquisition too, all three groups are also predicted to pass through stage B. However, whereas L2 adults should progress beyond stage B to stage C as their L2 proficiency increases, L1 and L2 children (within the relevant age range) will remain in stage B until their discourse integration

[17] The reader may recall that in the production experiment, a handful of subjects scrambled indefinite objects across negation where they were not supposed to, that is, in the condition designed to elicit a non-specific indefinite. However, as noted in §5.4.3.4, given that the non-scrambled order entails the scrambled order – that is, given that in the situation where the non-specific reading is true, the specific reading is also true – the scrambled order is not ungrammatical. It is possible therefore that the subjects intended a specific reading.

abilities increase with age. This means that in stage C, we should only find older L1 children, older, higher proficiency L2 children and higher proficiency L2 adults, or perhaps more accurately, given that there are some younger L1 and L2 children and low proficiency L2 adults with targetlike comprehension, we should predominantly find these groups at this stage.

SUMMARY

In this section, I outlined how the interpretive constraints on scrambled indefinite objects in Dutch might be acquired by L1 children and English-speaking L2 children and adults. I speculated that learners make use of a Blocking Principle, which means that once they associate the non-scrambled and scrambled forms by means of syntactic movement, the non-specific interpretation of the scrambled object is ruled out because for this interpretation, a less 'costly' alternative, the non-scrambled form, is available. The combination of the Blocking Principle with the discourse integration problems outlined in the previous section led to the postulation of three acquisitional stages for all three groups of learners.

7.4.2 An alternative account: De Hoop and Krämer (to appear)

As noted above, the notion of blocking is also employed by de Hoop and Krämer (to appear) as part of a recent OT analysis of the L1 acquisition of scrambling. De Hoop and Krämer propose that children's problems with the specific reading of the indefinite object in comprehension stem from their inability, as hearers, to take the speaker's perspective into account. Although the proposal put forward here is very much in the spirit of de Hoop and Krämer's, there are, I believe, several crucial differences. Before explaining what these differences are exactly, I first present the basics of de Hoop and Krämer's analysis. As a detailed presentation and analysis of OT is beyond the scope of this thesis, I outline only those aspects of the theory which are necessary to explain the general idea behind their proposal. The reader is referred to the original paper and the references therein for the specifics.

In OT (Prince and Smolensky 1997), constraints are used to evaluate a candidate set of outputs generated for a given input. In OT syntax, the input is generally a semantic structure and the output is a set of candidate forms, whereas in OT semantics, the input is a syntactic structure and the output a set of candidate meanings. Optimising from meaning to form and vice versa is unidirectional optimisation. The former takes the speaker's perspective ('Which form should I use to express this meaning?'), whereas the latter adopts the hearer's perspective ('Which meaning is expressed using this form?). Forms and meanings can also be evaluated as pairs, however. This is bidirectional optimisation (Blutner 2000). Bidirectional optimisation requires the speaker to take both her perspective and that of the hearer into account when producing an utterance, and likewise for the hearer, both his perspective and that of the speaker should be taken into account when interpreting an utterance. This can be paraphrased as follows: the hearer considers the alternative forms the speaker might have used ('If she had wanted to me to interpret it in this way, then she would have said that') and the speaker considers the alternative meanings which might be understood by the hearer ('If I say it like this, he might interpret it like that, whereas if I said it in a different way….'). De Hoop and Krämer's claim is that this type of reasoning, that is, bidirectional optimisation, is what children cannot do and this is why they fail to access the specific interpretation for the indefinite object.

Let us consider how this works for scrambling in more detail. There are two forms, scrambled and non-scrambled, and two meanings, specific and non-specific.[18] For indefinite object NPs, the scrambled form is marked (because it is less frequent) and the non-specific interpretation is the unmarked meaning (because it is of type <*e*,*t*>, the basic type of indefinite NPs, others being derived using type-shifting operations (see §2.3.3.1 for details)). There are thus four different combinations of <form, meaning> pairs, namely <scrambled, specific>, <scrambled, non-specific>, <non-scrambled, specific> and <non-scrambled, non-specific>. These <form, meaning> pairs are evaluated with respect to a number of ranked constraints on meaning (M1 and M2) and form (F1), which are taken from the literature:

(13) Constraints on meaning and form in de Hoop and Krämer's analysis

M1	Objects get a non-specific interpretation
M2	Indefinite NPs get a non-specific interpretation
F1	Indefinite objects do not scramble

These constraints are soft, that is, they are violable. When two constraints are in conflict with each other, it is their relative strength which determines the optimal candidate out of the set of candidates. In other words, violating a weaker constraint is less serious than violating a stronger constraint. A candidate is optimal if all other candidates show more serious violations. However, because constraints are violable, optimal candidates can sometimes violate constraints. The constraints in (13) are given in descending order of strength. For OT syntax, F1 is the relevant constraint: if the input is an indefinite object, the optimal expression of this would therefore be a non-scrambled form. In OT semantics, M1 and M2 apply: if the input is an indefinite object, it is optimally interpreted as non-specific. Thus, when unidirectional optimisation occurs, the optimal form for an indefinite object is in non-scrambled position and its optimal interpretation is non-specific, that is, the result is the unmarked meaning and the unmarked form. When bidirectional optimisation applies, two super-optimal pairs emerge, namely the unmarked form with the unmarked meaning, that is <non-scrambled, non-

[18] De Hoop and Krämer label the two readings 'referential' and 'non-referential'. The terms 'specific' and 'non-specific' are maintained here for clarity of exposition.

specific>, and the marked form with the marked meaning, that is <scrambled, specific>.[19] This is given in the OT tableau in (14), where * and ✓ designate a constraint violation and no violation, respectively. The super-optimal pairs are indicated with the ✌ symbol.

(14) Bidirectional OT tableau for indefinite objects
(adapted from de Hoop and Krämer to appear, ex. 22)

Indefinite object	M1	M2	F1
✌ <non-scrambled, non-specific>	✓	✓	✓
<non-scrambled, specific>	*	*	✓
<scrambled, non-specific>	✓	✓	*
✌ <scrambled, specific>	*	*	*

The other two <form, meaning> pairs are blocked because for each, there is another pair, namely the <non-scrambled, non-specific> pair, which has either a more optimal form or a more optimal meaning. The <non-scrambled, specific> pair is thus blocked by the <non-scrambled, non-specific> pair because the latter has a more optimal (i.e. non-specific) meaning for the indefinite object and therefore does not violate the constraints M1 and M2. Likewise, the <scrambled, non-specific> pair is also blocked by the <non-scrambled, non-specific> pair because the latter has a more optimal (i.e. non-scrambled) form and hence, does not incur a violation of constraint F1. This leaves the <scrambled, specific> pair, which is allowed because there is no other super-optimal pair which can block it. It is this process of blocking which is characteristic of bidirectional optimisation and which should be viewed – from a comprehension point of view – as taking the speaker's perspective into account. Essentially, what this means, for an adult, is that when he hears a scrambled indefinite object NP, he knows that given that the speaker used a sub-optimal (scrambled) form she must have intended a sub-optimal (specific) meaning, because if the optimal (non-specific) meaning were intended, she would have used the optimal (non-scrambled) form. According to the bidirectional OT account, this is the type of 'reasoning' which children cannot carry out and as a result, they assign the indefinite its optimal (non-specific) meaning (see Hendriks and Spenader (2004) for a similar approach to the L1 acquisition of Principle B and Hendriks (2005) on the L1 acquisition of stress shift and focus).

The proposal which I made above is thus similar to de Hoop and Krämer's approach in that both employ the notion of blocking to compare different <form, meaning> pairings. The role which blocking takes in the

[19] This is more or less Williams' observation regarding the more complex ends of the two related dimensions being associated with each other – except that here, it is embedded in a theoretical framework rather than simply being stated as a generalisation.

acquisition process on these two approaches is, I believe, quite different, however. There are two crucial differences. First, for de Hoop and Krämer, blocking is a consequence of learning how to optimise bidirectionally, whereas on my proposal, blocking is viewed as a principle of language acquisition which is given.[20] Second, for de Hoop and Krämer, blocking, that is, bidirectional optimisation, is acquired late (in comparison with other properties of language) as a consequence of children learning to take the speaker's perspective. As noted above, on my proposal, blocking is not acquired at all and for this reason, it does not necessarily have to be associated with delayed acquisition. Rather, I suggest that the observed delay in targetlike comprehension is not actually related to blocking at all.

The differences between these two approaches highlights an important problem which de Hoop and Krämer's analysis faces. This concerns the observed disparity between production and comprehension. On their analysis, L1 children's non-targetlike comprehension of scrambled indefinite objects is the result of unidirectional rather than bidirectional optimisation. Once bidirectional optimisation is available, children's comprehension should be targetlike. This also holds for production as well, however: once bidirectional optimisation is available, children's production should also be targetlike.[21] In other words, until this point in their development, L1 children should not use the marked (scrambled) form to express the marked (specific) meaning. The data presented here suggest that this is not the case, however, neither for L1 children nor for L2 children. It is possible though, that when these children scramble, the form they produce is not marked (H. de Hoop p.c., 12 May 2005). Rather, as we have observed for the native adults, the children may realise the scrambled object as the cardinal/stressed determiner *één* 'one'. This would not be a problem for de Hoop and Krämer's analysis. This is because – on the assumption that *één* N is a cardinal NP – the scrambled order and the specific meaning are both unmarked for cardinal NPs, where marked should be understood as most frequent.[22] As demonstrated above for indefinite NPs, the link between an unmarked form and an unmarked meaning is achieved using unidirectional optimisation only. If, therefore, when children scramble, they use a cardinal NP rather than an indefinite, this would not involve bidirectional optimisation and hence, de Hoop and Krämer's claim regarding the role of bidirectional optimisation would hold.

So what do L1 children do? The 13 L1 children reported on in Chapter 5 produced in total 34 scrambled objects in the specific indefinite condition.

[20] In order for bidirectional optimisation to apply successfully, it is essential that children know the constraints given in 0. It is not clear, however, where knowledge of these constraints should come from.

[21] Note that this is essentially the same prediction as the Blocking Principle makes.

[22] As noted in §2.1.2, it is unclear whether the *één* form is a cardinal or stressed indefinite determiner.

Of these 34 scrambled objects, 16 (47.1%) were realised as *één*. This means that, on the assumption that *één* is a cardinal determiner and not a stressed indefinite determiner, half of the scrambled objects produced by children in this condition are not 'true indefinites'.[23] The individual results are given in Table 61.

Table 61. L1 children: Distribution of objects realised as *één* in specific indefinite condition in production task

Subject	Specific indefinite condition		
	% scrambled	No. of objects realised as *één*	
		Neg-Obj-Verb	Obj-Neg-Verb
L1#1P	75% (3/4)	0/1	0/3
L1#2P	16.7% (1/6)	0/5	1/1
L1#3P	0% (0/6)	0/6	---
L1#4P	100% (4/4)	---	1/4
L1#5P	25% (1/4)	0/3	0/1
L1#6P	100% (4/4)	---	3/4
L1#7P	83.3% (5/6)	0/1	2/5
L1#8P	50% (1/2)	0/1	0/1
L1#9P	50% (2/4)	0/2	1/2
L1#10P	100% (6/6)	---	3/6
L1#11P	66.7% (2/3)	1/1	1/2
L1#12P	100% (4/4)	---	4/4
L1#13P	33.3% (1/3)	0/2	0/1
TOTAL	60.7% 34/56)	4.5% (1/22)	47.1% (16/34)

Of all the eight children who produce scrambled *één*, only one (L1#12P) uses this form exclusively. The others also use the indefinite determiner *een*. Excluding the tokens with *één*, the L1 children's rate of scrambling decreases slightly from 60.7% to 46.2% (18/39).[24] This proportion of targetlike responses in production still contrasts with the proportion of targetlike responses in comprehension in Krämer's original experiment, however: compare 16.7% (7/42) for the 4-5-year-olds and 22.9% (11/48) for the 5-6-year-olds. The use of *één* in the child L2 data is difficult to evaluate in this regard because of the interfering factors of L1 transfer (cf. §5.4.3.3; see Appendix C for data). Nevertheless, the L1 data alone suggest that 5-year-old children do produce scrambled indefinites (i.e. *een* only) in the relevant context, albeit not at

[23] Interestingly, the adult native controls produced tokens with *één* in 88.1% (37/42) of items.
[24] This figure is calculated as follows. The number of scrambled *één* tokens is subtracted from the total number of scrambled tokens (34-16=18) and this is divided by the total number of (scrambled and non-scrambled) tokens minus the number of tokens realised as *één* (56-17=39).

adultlike levels, and that the production ~ comprehension discrepancy remains. This goes against de Hoop and Krämer's proposal.[25]

It is of course possible that once the methodological concerns outlined above are addressed, and in particular, the issue of the felicity of negation, the difference between production and comprehension might disappear, in which case, the data would be in line with de Hoop and Krämer's analysis. If production and comprehension were observed to develop simultaneously, however, it would mean that these authors would have to assume that bidirectional optimisation is available (at least for some children) as early as age five. This, in turn, would undermine the link between the acquisition of scrambling and other 'late acquired' properties of language which have been claimed to follow from bidirectional optimisation, such as Principle B (see references above).

SUMMARY

In their OT analysis of the acquisition of scrambled indefinite objects, de Hoop and Krämer (to appear) also make use of the notion of blocking. For them, blocking is part of the process of bidirectional optimisation, which is linked to children's ability to take the speaker's perspective, an ability which they are assumed to acquire relatively late. De Hoop and Krämer's use of blocking was shown to differ in two important ways from how I speculated blocking might be used. Furthermore, their predictions regarding production and comprehension were not supported by the empirical findings currently available.

7.5 Summary

This chapter compared the production and comprehension of scrambled indefinite objects by L1 and L2 children. The data from 12 individual L2 children who took part in both experiments confirmed the observation made for the group results, namely that there is a discrepancy between production and comprehension, with targetlike comprehension being delayed with respect to targetlike production. It is possible that methodological factors may have contributed to this result, namely the potentially infelicitous use of negation and the salience of the various objects in the two truth value judgement tasks. Assuming that despite these methodological issues, the production ~ comprehension discrepancy were still to hold, I explored two possible explanations for this finding, and I concluded, following Krämer (2000) and Miller and Schmitt (2004), that limited discourse integration was the likely cause of children's problems on the comprehension side. Finally, I made a speculative proposal regarding how the interpretive constraints on scrambled

---

[25] If the *één* form is analysed as a stressed indefinite (Barbiers 2005), all the production data go against de Hoop and Krämer's analysis.

indefinites might be acquired, namely by means of a Blocking Principle (Williams 1997, amongst others). I hypothesised that once learners link the scrambled and non-scrambled orders by means of syntactic rule, the non-specific reading for the scrambled indefinite object is blocked because for this reading, a 'cheaper' option, the non-scrambled order, is available. This proposal was compared and contrasted with a recent OT analysis put forward by de Hoop and Krämer (to appear).

The discussion in this chapter would be further informed by new data. In particular, what is needed are the following sets of data: (i) a series of experiments which control for the salience of the different objects in the experimental scenario and the felicity of negation, and which use *één* instead of *een*; and (ii) more production data from L1 children (at least from 4-year-olds and 6-year-olds), preferably also for scrambling across other scopal operators besides negation (although the results in Schaeffer (2000) suggest that this might be difficult to elicit).

CHAPTER 8

SUMMARY AND IMPLICATIONS

Introduction

This chapter presents a summary of the findings and conclusions of this thesis. It also considers the implications of these findings for previous studies and then reviews some of the methodological issues which were raised in the course of this study. Finally, some final words are dedicated to areas for future research.

8.1 Summary of main findings

This thesis compared and contrasted three different groups of learners – L2 children, L2 adults and L1 children – in their acquisition of direct object scrambling in Dutch. The rationale behind the present study was a proposal put forward by Schwartz (1992). Schwartz claimed that child L2 acquisition can inform the debate about whether adult L2 acquisition is constrained by UG or whether it relies on general learning mechanisms (such as, for example, pattern-matching). Schwartz proposed that on the assumption (i) that child L2 acquisition is driven by UG and (ii) that because (L1 and L2) children make use of UG, the general learning mechanisms in question are more relevant to adult L2ers, then comparing the developmental sequences of child L2ers with those of adult L2ers, while holding the L1 constant, can provide evidence for or against UG involvement in adult L2 acquisition.

It was argued that for the L2 subjects in the present study, whose L1 was English, the acquisition of the interpretive constraints on scrambling presented a poverty-of-the-stimulus problem: knowledge that scrambled indefinite objects are restricted to a specific interpretation is not available in the L2 input, it cannot be transferred from the L1 (which does not have scrambling), and it is not the subject of classroom instruction. This was also claimed to hold (in a certain sense at least) for the acquisition of scrambled definite NPs in negative sentences. Therefore, it was concluded, if L2 adults and children demonstrate targetlike knowledge of the interpretive constraints on scrambling, their interlanguage grammars must somehow be constrained in the same way as L1 acquisition is.

The child L2 ~ adult L2 comparison was complicated by the observations made in previous studies that the comprehension of scrambled indefinites in L1 children remains non-targetlike at age seven and even older, much later than the age at which targetlike production of scrambled objects (at least of definite objects) is in place (Krämer 2000; Schaeffer 2000). The present study built on this previous research (i) by systematically investigating the production and comprehension of scrambling, including its interpretive effects, in L2 adults and L2 children, and (ii) by providing new child L1 data which rectified some of the methodological problems in previous studies and expanded the domain of enquiry to older children.

An elicited production experiment and two truth value judgement tasks were conducted using exactly the same methodology with all three groups of learners. The judgement data were also supplemented with an act-out task carried out with a subset of the L2 adults. For indefinite objects, the production experiment examined whether learners produced the targetlike form (scrambled vs. non-scrambled) in the relevant context (specific vs. non-specific), and for definite objects, it determined whether a scrambled form was produced in a context where sentential negation was intended. The comprehension tasks asked whether learners assigned the targetlike interpretation (specific vs. non-specific) to the relevant form (scrambled vs. non-scrambled). The first truth value judgement task tested learners' interpretation of indefinite objects scrambled across negation, as in (1), and the second judgement task and the act-out task tested scrambling across the frequency adverbial *twee keer* 'twice', as illustrated in (2).

(1)	De jongen	heeft	een vis	niet	gevangen
	the boy	has	a fish	not	caught
	'The boy didn't catch a (certain) fish.'				

(2)	a.	Het meisje	heeft	een aap	twee keer	gekieteld
		the girl	has	a monkey	two times	tickled
		'The girl tickled a (certain) monkey twice.'				
	b.	Het meisje	heeft	twee keer	een aap	gekieteld
		the girl	has	two times	a monkey	tickled
		'The girl tickled a(ny) monkey twice.'				

The main findings were as follows. In production, L2 children and L2 adults were found to pass through the same developmental sequence. This finding is compatible with Schwartz's (1992) proposal. The sequence for definite and specific indefinite objects is given in (3) and the sequence for non-specific indefinites in (4).

(3) **Developmental sequence for the acquisition of scrambled definite and specific indefinite objects in Dutch by English-speaking L2 children and L2 adults**

Stage	Description	Word order patterns produced
I	No scrambling	Negation-Verb-Object only
II-a	No scrambling	Negation-Object-Verb *and* Negation-Verb-Object
II-b	No scrambling	Negation-Object-Verb only
III-a	Optional scrambling	Object-Negation-Verb *and* Negation-Object-Verb
III-b	Scrambling	Object-Negation-Verb only

(4) **Developmental sequence for the acquisition of non-scrambled non-specific indefinite objects in Dutch by English-speaking L2 children and L2 adults**

Stage	Description	Word order patterns produced
I	Non-targetlike no scrambling	Negation-Verb-*een* NP only
II	Targetlike non-scrambled order	Negation-*een* NP-Verb only
III-a	Optional use of target lexical form	*geen* NP-Verb *and* Negation-*een* NP-Verb
III-b	Targetlike non-scrambled order and lexical form	*geen* NP-Verb

The mid and high proficiency children and adults scrambled definite and specific indefinite objects sometimes or always in the relevant contexts (i.e. definite and specific indefinite), and the high proficiency subjects generally always scrambled. The L2ers made the relevant distinction between specific and non-specific indefinites, that is, they scrambled the former while leaving the latter in the non-scrambled base position.

The data from the L1 5-year-olds indicated that they, too, made the relevant distinction between specific and non-specific indefinites. The developmental sequence for the L1 children for definite and specific indefinite objects, which were scrambled at comparable rates, is given in (5) and the sequence for non-specific indefinites in (6).

(5) **Developmental sequence for the acquisition of scrambled definite and specific indefinite objects in Dutch by L1 children**

Stage	Description	Word order patterns produced
I	No scrambling	Negation-Object-Verb only
II-a	Optional scrambling	Object-Negation-Verb *and* Negation-Object-Verb
II-b	Scrambling	Object-Negation-Verb only

(6) **Developmental sequence for the acquisition of non-scrambled non-specific indefinite objects in Dutch by L1 children**

Stage	Description	Word order patterns produced
I	Targetlike non-scrambled order	Negation-*een N*-Verb only
II-a	Optional use of target lexical form	*geen N*-Verb *and* Negation-*een N*-Verb
II-b	Targetlike non-scrambled order and form	*geen N*-Verb only

The only difference between the L1 and L2 stages was the presence of the Negation-Verb-Object stage in the L2 sequence. This was argued to result from L1 transfer.

In comprehension, consonant with the results of previous studies, L1 children were observed to assign a non-targetlike, non-specific interpretation to scrambled indefinite objects as old as age 13 on the negation task and age 11 on the *twee keer* task. From the age of nine, however, the child L1 groups did not significantly differ from the native adults. On the *twee keer* task, the child L2 data suggested the same age-related development as the L1 children: the younger children at time of testing were significantly different from the older children and natives, and unlike the high proficiency L2 adults, the high proficiency L2 children were significantly different from the native adults. On the act-out task, which was also on scrambling across *twee keer*, the high proficiency L2 adults also demonstrated targetlike knowledge of the same preferences as native adults, manipulating one object in the non-scrambled condition on significantly fewer occasions than in the scrambled condition. Both the truth value judgement task on *twee keer* results and the results of the act-out task contrast with those of the judgement task on scrambling across negation. On this task, the L2 children and adults consistently performed poorly. There was no development with age for the L2 children or with proficiency for either group. The L2ers' poorer performance on the negation task was attributed to a combination of the potentially infelicitous use of negation in the experimental scenario (Gualmini 2003) and L1 transfer.

The discrepancy between production and comprehension observed in earlier studies was also found in the present study, at the group level for both the L1 and the L2 children, and at the individual level, for the only subjects in common across tasks, namely a subset of the L2 children. For these 12 subjects, the following generalisation was made: subjects who (always or sometimes) had targetlike comprehension, (always or sometimes) had targetlike production. The opposite did not hold, however: subjects who (always or sometimes) had targetlike production did not (always or sometimes) have targetlike comprehension. Various explanations were explored for this finding. These included methodological issues, such as the salience of the objects in the experimental scenarios and the infelicitous use of negation, and following Krämer (2000) and Miller and Schmitt (2004), limited discourse integration.

The observation that there were English-speaking L2 children and adults who consistently rejected the non-specific reading for scrambled indefinite objects demonstrated that these learners were able to overcome the poverty-of-the-stimulus problem. This, in turn, suggested that L2 development makes use of the same mechanisms as L1 development. I speculated that in order to rule out the non-specific reading for the scrambled indefinite object, learners make use of a Blocking Principle (Williams 1997). The Blocking Principle ensures that once learners know that Dutch has the option of moving objects to a scrambled position, the non-specific reading for the scrambled indefinite is ruled out because for this reading, a 'cheaper' option, the non-scrambled order, exists.

8.2 Implications for previous studies

This section considers the possible implications of the findings of the present study for previous research on L1 and L2 acquisition.

8.2.1 L1 acquisition

The present study (partially) replicated two L1 studies, namely Schaeffer (2000) and Krämer (2000). The implications of the current findings for these two approaches is addressed in turn.

The first observation regarding Schaeffer's production experiment is that the systematic difference which the native adults made between the specific and non-specific conditions in the present study indicate that the modifications carried out to her original set-up were a success and that the lack of such a differentiation in her original experiment was the result of a methodological problem. The existence of an optional scrambling stage in the L2 data was claimed to be inconsistent with Schaeffer's approach. This was because, on her account, optional scrambling is tied to the lack of knowledge of a certain pragmatic principle. L1 children do not have this principle and hence scramble optionally. The L2 subjects tested here were old enough to know the relevant principle, however, and therefore, they should not have scrambled optionally; yet, they did. This finding could potentially be accounted for by L1 transfer (Schaeffer 2005); future work on L2ers of Dutch with different L1s is necessary to determine whether this is the case.

One of the goals of collecting new L1 data in the present study was to determine the age at which L1 children consistently accept a specific reading and reject a non-specific reading for scrambled indefinites. This turned out to be at around age 12, which was much later than the oldest age group which Krämer tested (7-year-olds). The results were nevertheless consistent with Krämer's Non-Integration Hypothesis, according to which the non-targetlike comprehension of scrambled indefinites stems from limited discourse integration abilities. The age-related effect in the child L2 data, coupled with the observation that unlike the high proficiency L2 adults, the high proficiency L2 children differed significantly from the native adults, suggested that Krämer's Non-Integration Hypothesis may hold for L2 acquisition, too. Further research is required, however, to pinpoint exactly which aspects of discourse integration are responsible for children's non-targetlike responses, and what, in turn, is the cause of children's problems in this domain. It was noted that Krämer's proposal was quite vague in this respect and that this made the formulation of clear falsifiable predictions rather difficult. Future work needs to address this issue in some detail, especially given that there is recent evidence suggesting that even young children have targetlike knowledge of discourse-related notions such as topic and focus and how these are instantiated in the adult grammar (e.g. De Cat 2002), and that under certain circumstances, young children are able to integrate discourse successfully (e.g. Song and Fisher 2005; Wijnen, Roeper and van der Meulen 2004).

8.2.2 L2 acquisition

The present study is one of the first to systematically compare L2 children, L2 adults and L1 children in their acquisition of one and the same phenomenon. By adhering as closely as possible to the desiderata given in Chapter 1, some of the methodological weaknesses observed in previous studies were avoided: the L1 was held constant across the two L2 groups, exactly the same methodology was employed across all three learner groups and an independent proficiency measure was used to make cross-group comparisons.

Where differences were observed, namely in the comprehension of scrambled indefinites, an independent explanation was available in limited discourse integration, which was assumed to be the source of the observed general age-effect. It was only as a result of the child L2 ~ child L1 comparison that such an effect could be detected. The present study thus demonstrated the crucial role which child L2 acquisition has to play in any L1/L2 comparison as well as the debate on the role of UG in adult L2 acquisition.

Previous studies on L2 scrambling focussed on the syntax of (clause-internal) scrambling only. By systematically investigating the interpretive properties of scrambling, the present study furnishes us with a more complete picture of how this particular property of Dutch develops in L2 acquisition. In particular, it also provides valuable data on the development of the syntax-semantics interface in child L2 acquisition. Consistent with previous studies on the acceptability of pre-subject scrambling in German (Hopp 2002; 2005; Schreiber and Sprouse 1998) and on the syntax-semantics interface in adult L2 acquisition (e.g. Dekydtspotter et al. 1999; Marsden 2004 amongst others), this study has demonstrated that L2 children and adults are able to overcome the poverty-of-the-stimulus.

8.3 Limitations of the present study

In the course of the study, certain limitations became apparent. In this section, I discuss three slightly more problematic aspects of the present study which should be borne in mind when interpreting the results and in future research in this area. These include potential problems with the experimental scenarios (as detailed in Chapters 6 and 7), the definition of adult L2ers, and the limitations of the proficiency measure. Each of these is discussed in turn.

In Chapter 6, I proposed that the use of negation in the truth value judgement task was potentially infelicitous. Following Gualmini (2003), I suggested that (at least some of) the non-targetlike responses on this task may have resulted from a failure in the experimental scenario to meet the felicity conditions for the use of negation. This effect appeared to be greater for the L2 subjects because it reinforced the non-targetlike response which L1 transfer would also lead them to. It is further possible that those objects most relevant to the non-specific interpretation were the most salient on both the negation and the *twee keer* task and this, too, may have contributed to the high number of non-targetlike responses provided by the older L1 children and by the L2ers.

Throughout this thesis, I have argued that because the English-speaking L2ers tested here demonstrated targetlike knowledge of the interpretive constraints on scrambled indefinite objects in Dutch, their interlanguage grammars must be constrained in the same way as L1 acquisition is. The finding that on the two judgement tasks, L1 children aged eight and older still produced non-targetlike responses indicated however that this acquisition process must also involve factors external to language narrowly construed. As noted above, this pattern of late L1 development somewhat complicated the child L2 ~ adult L2 comparison. In the present study, a child L2er was defined as 'a non-native acquirer whose initial exposure to the target language is between the ages of four and seven'. Consequently, any L2ers whose age at initial exposure was eight or older were assumed to be adults. The delay in targetlike comprehension in L1 acquisition meant that any L2ers who had been categorised as adults on the basis of their age at first exposure, but who were in the range of non-targetlike L1 children in terms of their age at time of testing, had to be analysed separately from older L2 adults. The results of this analysis indicated that the younger L2 adults patterned more like the L2 children than like the older L2 adults. A more straightforward child L2 ~ adult L2 comparison might only have included L2 children who were children at time of testing and L2 adults who were clearly adults at time of testing, say, for example, aged 16 or older. An even more straightforward comparison would perhaps focus on a target language property which was acquired by L1 children before the age of four.

Although in the context of the present study, an independent reason was available for why the younger adults should differ from the older adults, the observation that age at time of testing can affect subjects' responses raises the issue of where the dividing line between L2 children and L2 adults should fall, not only in terms of age at first exposure, as in the definition adopted here, but also with respect to age at time of testing. In order to address this issue, these two age-related variables (and ideally also length of exposure) would have to be systematically controlled for, and data would have to be available from a sufficient number of subjects from a variety of proficiency levels within each age group. The feasibility of any study attempting to do all this would crucially depend on locating a large enough group of L2ers with a continuing history of immigration into the target language country as both children and adults.

The proficiency measure developed in Chapter 4 formed an invaluable part of the present study. Although it was carefully designed so that it did not disadvantage the L2 children in favour of the L2 adults and so that it did not test the same construct as the experimental tasks, it is not without its limitations. First, it was not possible to assign subjects a proficiency score when they failed to complete the task. Second, calculating the final score was very time-consuming as it required transcribing and coding semi-spontaneous data for each subject. Third, although this was not a problem in the present study, the following limitation of the proficiency measure should be borne in mind for future work. Because the proficiency scores which subjects are assigned are based on the relative position in the group of their scores on each of the sub-measures, their final scores will also be relative. In other words, subjects' scores will depend on the range of scores in the group of L2ers as a whole. Consequently, it is imperative that the subjects in a sample span the full proficiency range.

8.4 Future research

The present study raises a number of interesting issues to be addressed in future research.

With respect to the acquisition of scrambling, the most pressing goal for future research will be to establish whether the observed findings for the comprehension tasks result from the methodological issues outlined in Chapters 6 and 7. A series of experiments should therefore be carried out to determine whether manipulating the following factors leads to more targetlike responses: (i) the (in)felicitous use of negation; (ii) the salience of the objects in the experimental scenario; (iii) the use of *één* instead of *een*; (iv) the targetlike response ('yes' instead of 'no') in the *twee keer* task; and (v) following Miller and Schmitt (2004), the naturalness of the set of objects. If, after controlling for each of these variables, it turns out that compared to production, targetlike comprehension of scrambled indefinites is still delayed for L1 and L2 children, further research will be needed to pinpoint the exact cause of this delay, including the nature of any hypothesised limitation in discourse integration abilities. Ideally, one would test both knowledge of scrambling and discourse integration in the same children in order to establish the extent of any link between the two. Furthermore, in order to make the comparison between the comprehension and production of scrambled objects complete, it would be desirable to test the production of scrambling across *twee keer* or some other adverb as well. Another possible avenue of enquiry would be to test L1 German children's acquisition of scrambled indefinites because in that language, the indefinite determiner, *ein*, does not know a distinction akin to the *een* ~ *één* alternation in Dutch.

In order to determine the generalisability of the findings of this thesis, more studies comparing L2 children with L2 adults (and preferably also with L1 children) are needed. These should investigate the acquisition of different phenomena and different L1/TL combinations. It is hoped that the desiderata listed in Chapter 1 will serve as guidelines for future research in this domain.

Bibliography

Aarssen, J. (1996). *Relating events in two languages. Acquisition of cohesive devices by Turkish-Dutch bilingual children at school age*. PhD dissertation. University of Tilburg.

Abusch, D. (1994). 'The scope of indefinites.' *Natural Language Semantics* 2:83-135.

Adamson, H.D. (1988). *Variation Theory and Second Language Acquisition*. Washington D.C.: Georgetown University Press.

Adger, D. (1994). 'Meaning, movement and economy.' In R. Aranivch, W. Byrne, S. Preuss and M. Senturia (eds.). *Proceedings of the West Coast Conference on Formal Linguistics 13*. Stanford: CSLI. pp. 451-466.

Adger, D. (1996). 'Economy and optionality: Interpretations of subjects in Italian.' *Probus* 8:117-135.

Appel, R. (1984). *Immigrant children learning Dutch. Sociolinguistic and psycholinguistic aspects of second-language acquisition*. Dordrecht: Foris.

Arlman-Rupp, A.J.L., van Niekerk de Haan, D. and van de Sandt-Koenderman, M. (1976). 'Brown's early stages: Some evidence from Dutch.' *Journal of Child Language* 3:267-274.

Armon-Lotem, S. (1998). 'Root Infinitives in child second language acquisition.' *McGill Working Papers in Linguistics* 13:11-17.

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Asher, J. and Price, B. (1967). 'The learning strategy of total physical response: Some age differences.' *Child Development* 38:1218-1227.

Bailey, N., Madden, C. and Krashen, S. (1974). 'Is there a 'natural sequence' in adult second language learning?' *Language Learning* 24:235-243.

Bailyn, J.F. (2002). 'Scrambling: The state of the art.' *GLOT Newsletter* March-April.

Bakx, J., Jetten, M. and Korebits, L. (1999). *Grammatica in Gebruik. Nederlands voor Anderstaligen [Grammar in Use. Dutch for Speakers of Other Languages]*. Amsterdam: Intertaal.

Barbier, I. (2000). 'An experimental study of scrambling and object shift in the acquisition of Dutch.' In S. Powers and C. Hamann (eds.). *The Acquisition of Scrambling and Cliticization*. Dordrecht: Kluwer. pp. 41-69.

Barbiers, S. (2005). 'Variation in the morphosyntax of one.' *Journal of Comparative Germanic Linguistics* 8:159-183.

Becker, A. (2005). 'The semantic knowledge base for the acquisition of negation and the acquisition of finiteness.' In H. Hendriks (ed.). *The Structure of Learner Varieties*. Berlin: Mouton de Gruyter. pp. 263-314.

Belletti, A. and Hamann, C. (2000). 'Ça on fait pas! On the L2-Acquisition of French by two young children with different source languages.' In S.C. Howell, S.A. Fish and T. Keith-Lucas (eds.). *Proceedings of the 24th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 116-127.

Bennis, H. and Hoekstra, T. (1984). 'Gaps and parasitic gaps.' *The Linguistic Review* 4:29-87.

Bergsma-Klein, W. (1996). *Specificity in Child Dutch: An Experimental Study*. Unpublished MA thesis. Utrecht University.

Berman, R.A. and Slobin, D.I. (1986). *Coding Manual: Temporality in Discourse*. Berkeley: University of California, Institute of Cognitive Studies.

Bialystok, E. and Hakuta, K. (1994). *In Other Words: The Science and Psychology of Second Language Acquisition*. New York: Basic Books.

Bialystok, E. (1997). 'The structure of age: In search of barriers to second language acquisition.' *Second Language Research* 13:116-137.

Bialystok, E. and Miller, B. (1999). 'The problem of age in second-language acquisition: Influences from language, structure and task.' *Bilingualism: Language and Cognition* 2:127-145.

Birdsong, D. (1992). 'Ultimate attainment in second language acquisition.' *Language* 68:706-755.

Birdsong, D. (1999). 'Introduction: Whys and why nots of the Critical Period Hypothesis for second language acquisition.' In D. Birdsong (ed.). *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Lawrence Erlbaum. pp. 1-22.

Birdsong, D. and Molis, M. (2001). 'On the evidence for maturational constraints in second-language acquisition.' *Journal of Memory and Language* 44:235-249.

Blake, J., Quartaro, G. and Onorati, S. (1993). 'Evaluating quantitative measures of grammatical complexity in spontaneous speech samples.' *Journal of Child Language* 20:139-152.

Bley-Vroman, R. (1989). 'What is the logical problem of foreign language learning?' In S.M. Gass and J. Schachter (eds.). *Linguistic Perspectives on Second Language Acquisition*. Cambridge: Cambridge University Press. pp. 42-68.

Bley-Vroman, R. and Chaudron, C. (1994). 'Elicited imitation as a measure of second-language competence.' In E.E. Tarone, S.M. Gass and A.D. Cohen (eds.). *Research Methods in Second-Language Acquisition*. Hillsdale, New Jersey: Lawrence Erlbaum. pp. 245-261.

Blom, E. (2003). *From Root Infinitive to Finite Sentence. The Acquisition of Verbal Inflections and Auxiliaries*. PhD dissertation. Utrecht University.

Blom, E. (2005). Successive bilingualism within the critical period: Child L2 vs. child L1 and adult L2 acquisition. Paper presented at IASCL, Berlin, July 27 2005.

Blom, E. and Polišenská, D. (to appear). 'Verbal inflection and verb placement in first and second language acquisition.' In M. Vliegen (ed.). *Proceedings of the 29th Linguistics Colloquium*. Amsterdam: Peter Lang.

Blum-Kulka, S. and Olshtain, E. (1986). 'Too many words: Length of utterance and pragmatic failure.' *Studies in Second Language Acquisition* 8:165-179.

Blutner, R. (2000). 'Some aspects of optimality in natural language interpretation.' *Journal of Semantics* 17:189-216.

Bobaljik, J.D. (1995). *Morphosyntax: The Syntax of Verbal Inflection*. PhD dissertation. MIT.

Boeschouten, H. and Verhoeven, L. (1986). 'First language acquisition in a second language submersion environment.' *Applied Psycholinguistics* 7:241-255.

Bol, G.W. and Kuiken, F. (1988). *Grammaticale analyse van taalontwikkelingsstoornissen [Grammatical analysis of language development disorders]*. PhD dissertation. University of Amsterdam.

Bongaerts, T. (1999). 'Ultimate attainment in foreign language pronunciation: The case of very advanced late foreign language learners.' In D. Birdsong (ed.). *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Erlbaum. pp. 133-159.

Bos, P. (1997). *Development of bilingualism. A study of school-age Moroccan children in the Netherlands*. PhD dissertation. University of Tilburg.

Broeder, P., Extra, G. and van Hout, R. (1993). 'Richness and variety in the developing lexicon.' In C. Perdue (ed.). *Adult language acquisition: Cross-linguistic perspectives*. Cambridge: Cambridge University Press. pp. 145-163.

Brown, R. (1973). *A First Language*. Cambridge, MA: Harvard University Press.

Bruening, B. (2001). 'QR obeys superiority: Frozen scope and ACD.' *Linguistic Inquiry* 32:233-273.

Cancino, H., Rosansky, E.J. and Schumann, J.H. (1978). 'The acquisition of English negatives and interrogatives by native Spanish speakers.' In E.M. Hatch (ed.). *Second Language Acquisition. A Book of Readings*. Rowley, MA: Newbury House. pp. 207-230.

Cantone, K.F. and Müller, N. (2004). Code-switching at the interface of language-specific lexicons and the computational system. Ms. University of Hamburg.

Chabon, S.S., Kent-Udolf, L. and Egolf, D.B. (1982). 'The temporal reliability of Brown's mean length of utterance (MLU-m) measure with post-stage V children.' *Journal of Speech and Hearing Research* 25:117-124.

Chapman, R.S. (1981). 'Computing mean length of utterance in morphemes.' In J.F. Miller (ed.). *Assessing language production in children*. London: Edward Arnold. pp. 22-25.

Chaudron, C. and Parker, K. (1990). 'Discourse markedness and structural markedness. The acquisition of English noun phrases.' *Studies in Second Language Acquisition* 12:43-64.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1991). 'Some notes on economy of derivation and representation.' In R. Freidin (ed.). *Principles and Parameters in Comparative Grammar*. Cambridge, MA: MIT Press. pp. 417-454.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Cinque, G. (1993). 'A null theory of phrasal and compound stress.' *Linguistic Inquiry* 24:239-298.

Clahsen, H. (1984). 'The acquisition of German word order: A test case for a cognitive approach to L2 development.' In R. Andersen (ed.). *Second Language: A Cross-linguistic Perspective*. Rowley, MA: Newbury House. pp. 219-242.

Clahsen, H. and Muysken, P. (1986). 'The availability of Universal Grammar to adult and child learners - a study of the acquisition of German word order.' *Second Language Research* 2:93-119.

Clahsen, H. (1988). 'Critical phases of grammar development. A study of the acquisition of negation in children and adults.' In P. Jordens and J. Lalleman (eds.). *Language Development*. Dordrecht: Foris. pp. 123-148.

Clahsen, H. and Muysken, P. (1989). 'The UG Paradox in L2 Acquisition.' *Second Language Research* 5:1-29.

Clahsen, H. (1990/1991). 'Constraints on parameter setting: A grammatical analysis of some acquisition stages in German child language.' *Language Acquisition* 1:361-391.

Clahsen, H., Eisenbeiss, S. and Penke, M. (1996). 'Lexical learning in early syntactic development.' In H. Clahsen (ed.). *Generative Perspectives on Language Acquisition*. Amsterdam: John Benjamins. pp. 129-160.

Clark, H.H. (1977). 'Bridging.' In P.N. Johnson-Laird and P.C. Wason (eds.). *Thinking. Readings in Cognitive Science.* Cambridge: Cambridge University Press. pp. 411-420.

Collins, C. and Thráinsson, H. (1996). 'VP-internal structure and object shift in Icelandic.' *Linguistic Inquiry* 27:391-444.

Coopmans, P., Krul, M., Planting, E., Vlasveld, I. and van Zoelen, A. (2004). 'Dissolving a Dutch delay in the acquisition of syntactic and logophoric reflexives.' In A. Brugos, L. Micciulla and C. Smith, E. (eds.). *Proceedings of the 28th Boston University Conference on Language Acquisition.* Somerville, MA: Cascadilla Press. pp. 108-119.

Coppetiers, R. (1987). 'Competence differences between native and near-native speakers.' *Language* 63:253-280.

Corver, N. and van Riemsdijk, H. (1994a). 'Introduction: Approaches to and properties of scrambling.' In N. Corver and H. van Riemsdijk (eds.). *Studies on Scrambling: Movement and Non-Movement Approaches to Free Word-Order Phenomena.* New York: Mouton de Gruyter. pp. 1-15.

Corver, N. and van Riemsdijk, H. (eds.). (1994b). *Studies on Scrambling: Movement and Non-Movement Approaches to Free Word-Order Phenomena.* New York: Mouton de Gruyter.

Crain, S. and Thornton, R. (1998). *Investigations in Universal Grammar. A Guide to Experiments on the Acquisition of Syntax and Semantics.* Cambridge, MA: MIT Press.

Crystal, D. (1974). 'Review of Brown (1973).' *Journal of Child Language* 1:289-307.

Crystal, D., Fletcher, P. and Garman, M. (1976). *The grammatical analysis of language disability: A procedure for assessment and remediation.* New York: Elsevier.

Curtiss, S. and Yamada, J. (1985). *Comprehensive Language Evaluation - Receptive Measures (CYCLE-R) and the Comprehensive Language Evaluation - Elicited Production (CYCLE-E).*

De Cat, C. (2002). *French Dislocation.* Unpublished PhD dissertation. University of York.

de Haan, A. (1996). *De Verwerving van Morfologische Finietheid in het Nederlands [The Acquisition of Morphological Finiteness in Dutch].* MA thesis. University of Groningen.

de Hoop, H. (1992). *Case Configuration and Noun Phrase Interpretation.* Phd dissertation, published (1996) New York: Garland. University of Groningen.

de Hoop, H. (2000). 'Optional scrambling and interpretation.' In H. Bennis, M. Everaert and E. Reuland (eds.). *Interface Strategies.* Amsterdam: Royal Netherlands Academy of Arts and Sciences. pp. 153-168.

de Hoop, H. (2003). 'Scrambling in Dutch: Optionality and Optimality.' In S. Karmini (ed.). *Word order and Scrambling.* Oxford: Blackwell. pp. 201-216.

de Hoop, H., Haverkort, M. and van den Noort, M. (2004). 'Variation in form versus variation in meaning.' *Lingua* 114:1071-1089.

de Hoop, H. and Krämer, I. (to appear). 'Children's optimal interpretations of indefinite subjects and objects.' *Language Acquisition.*

De Jong, J.H.A.L. and Verhoeven, L. (1992). *The Construct of Language Proficiency.* Amsterdam: John Benjamins.

De Villiers, J.G. and Tager-Flusberg, H., B. (1975). 'Some facts one simply cannot deny.' *Journal of Child Language* 2:279-286.

De Villiers, J.G. (2005). The interfaces of language and Theory of Mind. Paper presented at workshop on "Language Acquisition between Sentence and Discourse", Radboud University Nijmegen, The Netherlands [13th May].

De Villiers, P. and De Villiers, J.G. (1973). 'A cross-sectional study of the development of grammatical morphemes in child speech.' *Journal of Psycholinguistic Research* 2:267-278.

Deen, K.U. (2002). *The Acquisition of Nairobi Swahili: The Morphosyntax of Inflectional Prefixes and Subjects*. Unpublished PhD dissertation. UCLA.

DeKeyser, R.M. (2000). 'The robustness of critical period effects in second language acquisition.' *Studies in Second Language Acquisition* 22:499-533.

Dekydtspotter, L., Sprouse, R.A. and Anderson, B. (1997). 'The interpretive interface in L2 acquisition: The process-result distinction in English-French interlanguage grammars.' *Language Acquisition* 6:297-332.

Dekydtspotter, L., Sprouse, R.A. and Anderson, B. (1998). 'Interlanguage A-bar dependencies: Binding construals, null prepositions and Universal Grammar.' *Second Language Research* 14:341-358.

Dekydtspotter, L., Sprouse, R.A., Swanson, K.A. and Thyre, R. (1999). 'Semantics, pragmatics and second language acquisition: The case of *Combien* extractions.' In A. Greenhill, H. Littlefield and C. Tano (eds.). *Proceedings of the 23rd Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 162-171.

Dekydtspotter, L., Sprouse, R.A. and Thyre, R. (1999/2000). 'The interpretation of quantification at a distance in English-French interlanguage: Domain specificity and second-language acquisition.' *Language Acquisition* 8:265-320.

Deprez, V. and Pierce, A. (1994). 'Crosslinguistic evidence for functional projections in early child grammar.' In T. Hoekstra and B.D. Schwartz (eds.). *Language Acquisition Studies in Generative Grammar.* Amsterdam: John Benjamins. pp. 57-84.

Dewaele, J.-M. (2000). 'Saisir l'insaisissable? Les mesures de longueur d'énoncés en linguistique appliquée [Seize the unseizable? Measures of Utterance Length and Applied Linguistics].' *International Review of Applied Linguistics* 38:17-33.

Diesing, M. (1992). *Indefinites.* Cambridge, MA: MIT Press.

Diesing, M. (1997). 'Yiddish VP order and the typology of object movement in Germanic.' *Natural Language and Linguistic Theory* 15:369-427.

Dimroth, C. (2005). Age effects on the process of L2 acquisition? Evidence from the acquisition of negation and finiteness in L2 German. Ms. Radboud University Nijmegen, The Netherlands.

Dulay, H. and Burt, M. (1973). 'Should we teach children syntax?' *Language Learning* 23:245-258.

Dulay, H. and Burt, M. (1974). 'Natural sequences in child second language acquisition.' *Language Learning* 24:37-53.

duPlessis, J., Solin, D., Travis, L. and White, L. (1987). 'UG or not UG, that is the question: A reply to Clahsen and Muysken.' *Second Language Research* 3:56-75.

Eisenbeiss, S. (1994). 'Kasus und Wortstellungsvariation im deutschen Mittelfeld. Theoretische Überlegungen und Untersuchungen zum Erstspracherwerb [Case and Word Order Variation in the German 'Mittelfeld'. Theoretical Considerations and Investigations in First Language Acquisition].' In B. Haftka (ed.). *Was*

Determiniert Wortstellungsvariation? [What Determines Word Order Variation?]. Opladen: Westdeutscher Verlag. pp. 279-298.

Emmorey, K. (2002). 'The Critical Period Hypothesis and the effects of late language acquisition.' In K. Emmorey (ed.). *Language, Cognition and the Brain.* Mahwah, NJ: Lawrence Erlbaum Associates. pp. 205-226.

Enç, M. (1991). 'The semantics of specificity.' *Linguistic Inquiry* 22:1-25.

Ervin-Tripp, S.M. (1978). 'Is second language learning like the first?' In E.M. Hatch (ed.). *Second Language Acquisition. A Book of Readings.* Rowley, MA: Newbury House. pp. 190-206.

Felix, S.W. (1977). 'Interference, interlanguage and related issues.' In C. Molony, H. Zobl and W. Stölting (eds.). *Deutsch im Kontakt mit Anderen Sprachen [German in Contact with Other Languages].* Kronberg: Scriptor. pp. 237-258.

Felix, S.W. (1985). 'More evidence on competing cognitive systems.' *Second Language Research* 1:47-72.

Fodor, J.D. and Sag, I.A. (1982). 'Referential and quantificational indefinites.' *Linguistics and Philosophy* 5:355-398.

Fromkin, V., Krashen, S., Curtiss, S., Rigler, D. and Rigler, M. (1974). 'The development of language in Genie: A case of language acquisition beyond the "critical period".' *Brain and Language* 1:87-107.

Gavruseva, E. and Lardiere, D. (1996). 'The emergence of extended phrase structure.' In A. Stringfellow, D. Cahana-Amitay, E. Hughes and A. Zukowski (eds.). *Proceedings of the 20th Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 225-236.

Gavruseva, E. (1998). 'Left-branch violations in child L2 English.' In A. Greenhill, M. Hughes, H. Littlefield and H. Walsh (eds.). *Proceedings of the 22nd Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 235-245.

Gavruseva, E. (2000). 'Aspect parameter in the guise of Optional Infinitives in child L2 English.' In S.C. Howell, S.A. Fish and T. Keith-Lucas (eds.). *Proceedings of the 24th Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 319-330.

Geurts, B. (2002). 'Specific indefinites, presupposition and scope.' In R. Bäuerle, U. Reyle and T.E. Zimmermann (eds.). *Presuppositions and Discourse.* Oxford: Elsevier.

Gilkerson, J. (2004). *Acquiring English particle verbs: Age and transfer effects in L2 acquisition.* Paper presented at the inaugural Generative Approaches to Language Acquisition North America (GALANA), University of Hawai'i, Honolulu, USA. [17th December].

Goodluck, H. (1986). 'Language acquisition and linguistic theory.' In P. Fletcher and M. Garman (eds.). *Language Acquisition.* Cambridge: Cambridge University Press. pp. 49-68.

Greenhouse, J.B. and Iyengar, S. (1994). 'Sensitivity analysis and diagnostics.' In H. Cooper and L.V. Hedges (eds.). *The Handbook of Research Synthesis.* New York: Russell Sage Foundation. pp. 383-398.

Gregg, K.R. (1996). 'The logical and developmental problems of second language acquisition.' In W.C. Ritchie and T.K. Bhatia (eds.). *Handbook of Second Language Acquisition.* San Diego: Academic Press. pp. 49-81.

Grewendorf, G. and Sternefeld, W. (1990). 'Scrambling theories.' In G. Grewendorf and W. Sternefeld (eds.). *Scrambling and Barriers.* Amsterdam: John Benjamins. pp. 3-37.

Grewendorf, G. and Sabel, J. (1999). 'Scrambling in German and Japanese: Adjunction versus multiple specifiers.' *Natural Language and Linguistic Theory* 17:1-65.

Grondin, N. and White, L. (1996). 'Functional categories in child L2 acquisition of French.' *Language Acquisition* 5:1-34.

Grosjean, F. (1998). 'Studying bilinguals: Methodological and conceptual issues.' *Bilingualism: Language and Cognition* 1:131-149.

Gualmini, A. (2003). '*Some* knowledge children don't lack.' In B. Beachley, A. Brown and F. Conlin (eds.). *Proceedings of the 27th Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 276-287.

Guasti, M.-T. (2002). *The Growth of Grammar.* Cambridge, MA: MIT Press.

Guéron, J. (1980). 'On the syntax and semantics of PP extraposition.' *Linguistic Inquiry* 11:637-678.

Haberzettl, S. (1999). '*Katze Maus essen* vs. *Katze essen Maus*: Die L1 als Königs- oder Holzwegbereiter zur L2? Zum Einfluss des L1-Wissens im Erwerb der deutschen Verbstellung durch türkische und russische Kinder.' In H.O. Spillmann and I. Warnke (eds.). *Internationale Tendenzen der Syntaktik, Semantik und Pragmatik [International Trends in Syntax, Semantics and Pragmatics].* Frankfurt: Peter Lang. pp. 157-165.

Haegeman, L. (1995). 'Root infinitives, tense and truncated structures in Dutch.' *Language Acquisition* 4:205-255.

Haeseryn, W. (1997). *Algemene Nederlandse Spraakkunst.* Groningen: Martinus Nijhoff.

Haider, H. and Rosengren, I. (1998). *Scrambling.* Lund: University of Lund.

Haider, H. and Rosengren, I. (2003). 'Scrambling: Nontriggered chain formation in OV languages.' *Journal of Germanic Linguistics* 15:203-267.

Haider, H. (2004). 'Pre- and postverbal adverbials in VO and OV.' *Lingua* 114:779-807.

Hakuta, K. (1976). 'A case study of a Japanese child learning English as a second language.' *Language Learning* 26:321-351.

Ham, E., Tersteeg, W.H.T.M. and Zijlmans, L. (1998). *Help! Een Cursus voor Buitenlanders (Deel 1/2) [Help! A Course for Foreigners. (Part 1/2)].* Utrecht: NCB.

Harley, B. and Wang, W. (1997). 'The Critical Period Hypothesis: Where are we now?' In A.M.B. de Groot and J.F. Kroll (eds.). *Tutorials in Bilingualism: Psycholinguistic Perspectives.* London: Lawrence Erlbaum. pp. 19-52.

Haznedar, B. (1997). 'L2 acquisition by a Turkish-speaking child: Evidence of L1 influence.' In E. Hughes, M. Hughes and A. Greenhill (eds.). *Proceedings of the 21st Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 245-256.

Haznedar, B. and Schwartz, B.D. (1997). 'Are there Optional Infinitives in child L2 acquisition?' In E. Hughes, M. Hughes and A. Greenhill (eds.). *Proceedings of the 21st Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 257-268.

Haznedar, B. (2001). 'The acquisition of the IP system in child L2 English.' *Studies in Second Language Acquisition* 23:1-39.

Haznedar, B. (2003). 'The status of functional categories in child second language acquisition: Evidence from the acquisition of CP.' *Second Language Research* 19:1-41.

Hendriks, P. and Spenader, J. (2004). When production precedes comprehension: An optimization approach to the acquisition of pronouns. Ms. University of Groningen.

Hendriks, P. (2005). Production/comprehension asymmetries and bidirectional Optimality Theory. Paper presented at workshop on Language Acquisition between Sentence and Discourse, Radboud University Nijmegen, The Netherlands [13th May].

Henry, A. and Tangney, D. (1996). 'Functional categories in child L2 acquisition of Irish.' In A. Stringfellow, D. Cahana-Amitay and A. Zukowski (eds.). *Proceedings of the 20th Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 318-127.

Hickey, T. (1991). 'Mean length of utterance and the acquisition of Irish.' *Journal of Child Language* 18:553-569.

Hoekstra, T. and Jordens, P. (1994). 'From adjunct to head.' In T. Hoekstra and B.D. Schwartz (eds.). *Language Acquisition: Studies in Generative Grammar: Papers in Honor of Kenneth Wexler from the 1991 GLOW Workshops.* Amsterdam: John Benjamins. pp. 119-149.

Hoekstra, T. and Hyams, N. (1996). 'Missing heads in child language.' In C. Koster and F. Wijnen (eds.). *Proceedings of Generative Approaches to Language Acquisition 1995.* Groningen: Centre for Language and Cognition. pp. 251-260.

Holmberg, A. (1986). *Word order and syntactic features.* PhD dissertation. University of Stockholm.

Hopp, H. (2002). *Constraints on Word Order Variation: Learnability and UG in Advanced English-German and Japanese-German Interlanguage.* Unpublished MA thesis. University of Durham.

Hopp, H. (2003). Scrambling in German and Dutch - An Interface-Economic Approach. Ms. Univeristy of Groningen.

Hopp, H. (2005). 'Constraining second language word order optionality: Scrambling in advanced English-German and Japanese-German interlanguage.' *Second Language Research* 21:34-71.

Hornstein, N. (1995). *LF: The Grammar of Logical Form from GB to Minimalism.* Oxford: Blackwell.

Hunt, K.W. (1970). *Syntactic Maturity in Schoolchildren and Adults.* Chicago: University of Chicago Press.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. and Trueswell, J.C. (2000). 'One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension.' *Journal of Psycholinguistic Research* 29:597-626.

Hyltenstam, K. and Abrahamsson, N. (2003). 'Maturational constraints in SLA.' In C. Doughty and M.H. Long (eds.). *The Handbook of Second Language Acquisition.* Oxford: Blackwell. pp. 539-588.

Ionin, T. and Wexler, K. (2002). 'Why is 'is' easier than '-s?: Acquisition of tense/agreement morphology by child second language learners of English.' *Second Language Research* 18:95-136.

Ioup, G., Boustagui, E., El Tigi, M. and Moselle, M. (1994). 'Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment.' *Studies in Second Language Acquisition* 16:73-98.

Iwasaki, N. (2003). 'L2 acquisition of Japanese: Knowledge and use of case particles in SOV and OSV sentences.' In S. Karmini (ed.). *Word Order and Scrambling.* Oxford: Blackwell. pp. 273-300.

Jin, J. (2003). 'Acquisition of L2 English DP by Korean children.' *Reading Working Papers in Linguistics* 7:77-102.

Johnson, J.S. and Newport, E.L. (1989). 'Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language.' *Cognitive Psychology* 21:60-99.

Johnson, J.S. and Newport, E.L. (1991). 'Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language.' *Cognition* 39:215-258.

Johnson, K. (1991). 'Object positions.' *Natural Language and Linguistic Theory* 9:577-636.

Johnston, J.R. (2001). 'An alternate MLU calculation: Magnitude and variability of effects.' *Journal of Speech, Language and Hearing Research* 44:156-164.

Jordens, P. (1988). 'The acquisition of word order in Dutch and German as L1 and L2.' *Second Language Research* 4:41-65.

Juffs, A. (1996). 'Semantics-syntax correspondences in second language acquisition.' *Second Language Research* 12:177-221.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic.* Dordrecht: Kluwer.

Kang, B. (2005). 'A learnability puzzle in scrambling.' In A. Brugos, M. Clark-Cotton, R. and S. Ha (eds.). *Proceedings of the 29th Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 331-340.

Karmiloff-Smith, A. (1979). *A Functional Approach to Child Language: A Study of Determiners and Reference.* Cambridge: Cambridge University Press.

Kayne, R. (1994). *The Antisymmetry of Syntax.* Cambridge, MA.: MIT Press.

Kellerman, E. (1995). 'Age before beauty: Johnson and Newport revisited.' In L. Eubank, L. Selinker and M. Sharwood-Smith (eds.). *The Current State of Interlanguage: Studies in Honor of William E. Rutherford.* Amsterdam: John Benjamins. pp. 219-232.

Kiparsky, P. (1973). '"Elsewhere" in phonology.' In S. Anderson and P. Kiparsky (eds.). *A Festschrift for Morris Halle.* New York: Holt, Rinehart and Winston. pp. 93-106.

Klee, T. and Fitzgerald, M.D. (1985). 'The relation between grammatical development and mean length of utterance in morphemes.' *Journal of Child Language* 12:251-269.

Koizumi, M. (1993). 'Object agreement phrases and the split VP hypothesis.' In C. Phillips and J.D. Bobaljik (eds.). *Papers on Case and Agreement I: MIT Working Papers in Linguistics.* Cambridge, MA: MIT, Department of Linguistics and Philosopy. pp. 99-148.

Koster, J. (1975). 'Dutch as an SOV language.' *Linguistic Analysis* 1:111-136.

Koster, J. (1999). 'The word order of English and Dutch: Collective and individual checking.' In W. Abraham (ed.). *Groninger Arbeiten zur germanistischen Linguistik.* Groningen: University of Groningen. pp. 1-42.

Koster, J. (2000). 'Pied Piping and the word orders of English and Dutch.' In M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim (eds.). *NELS 30: Proceedings of the North East Linguistic Society*. Amherst, MA: GLSA. pp. 415-426.

Krämer, I. (1993). 'The licensing of subjects in early child language.' *MIT Working Papers in Linguistics* 19:197-212.

Krämer, I. (2000). *Interpreting Indefinites. An Experimental Study of Children's Language Comprehension*. PhD dissertation. Utrecht University.

Krashen, S. (1973). 'Lateralization, language learning and the critical period: Some new evidence.' *Language Learning* 23:63-74.

Krashen, S., Long, M.H. and Scarcella, R. (1979). 'Age, rate and eventual attainment in second language acquisition.' *TESOL Quarterly* 13:573-582.

Krashen, S. (1982). 'Accounting for child-adult differences in second language rate and attainment.' In S. Krashen, R. Scarcella and M.H. Long (eds.). *Child-Adult Differences in Second Language Acquisition*. Rowley, MA: Newbury House. pp. 202-206.

Kroffke, S. and Rothweiler, M. (to appear). 'Variation im frühen Zweitspracherwerb des Deutschen durch Kinder mit türkischer Erstsprache.' In M. Vliegen (ed.). *Proceedings of the 39th Linguistics Colloquium*. Amsterdam: Peter Lang.

Lakshmanan, U. (1995). 'Child second language acquisition of syntax.' *Studies in Second Language Acquisition* 17:301-329.

Lakshmanan, U. and Ozeki, M. (1996). 'The case of the missing particle: Objective case assignment and scrambling in the early grammar of Japanese.' In A. Stringfellow, D. Cahana-Amitay, E. Hughes and A. Zukowski (eds.). *Proceedings of the 20th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 431-442.

Lalleman, J. (1986). *Dutch Language Proficiency of Turkish Children Born in The Netherlands*. Dordrecht: Foris.

Landis, J.R. and Koch, G.G. (1977). 'The measurement of observer agreement for categorical data.' *Biometrics* 33:159-174.

Lane, H. (1976). *The Wild Boy of Aveyron*. Cambridge, MA: Harvard University Press.

Larsen-Freeman, D.E. and Strom, V. (1977). 'The construction of a second language acquisition index of development.' *Language Learning* 27:123-134.

Larsen-Freeman, D.E. (1978a). 'An ESL index of development.' *TESOL Quarterly* 4:439-448.

Larsen-Freeman, D.E. (1978b). 'Evidence of the need for a second language acquisition index of development.' In W.C. Ritchie (ed.). *Second Language Acquisition Research: Issues and Implications*. New York: Academic Press. pp. 127-136.

Larsen-Freeman, D.E. (1983). 'Assessing global second language proficiency.' In H.W. Seliger and M.H. Long (eds.). *Classroom Oriented Research in Second Language Acquisition*. Rowley, MA: Newbury House. pp. 287-304.

Larsen-Freeman, D.E. and Long, M.H. (1991). *An Introduction to Second Language Acquisition Research*. London: Longman.

Lenneberg, E. (1967). *Biological Foundations of Language*. New York: Wiley & Sons.

Lidz, J. and Musolino, J. (2002). 'Children's command of quantification.' *Cognition* 84:113-154.

Lidz, J. and Musolino, J. (submitted). 'Continuity in linguistic development: The scope of syntax and the syntax of scope.'

Lightbown, P. (1977). *Consistency and Variation in the Acquisition of French: A Study of First and Second Language Development*. Unpublished PhD dissertation. Concordia University.

Loban, W. (1976). *Language Development: Kindergarten Through Grade Twelve*. Urbana, Illinois: National Council of Teachers of English.

Long, M.H. (1990). 'Maturational constraints on language development.' *Studies in Second Language Acquisition* 12:251-285.

Lust, B., Flynn, S. and Foley, C. (1996). 'What children know about what they say: Elicited imitation as a research method for assessing children's syntax.' In H. Smith Cairns (ed.). *Methods for Assessing Children's Syntax*. Cambridge, MA: MIT Press. pp. 55-76.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Marcus, G.F., Pinker, S., Ullman, M., Hollander, M., Rosen, T.J. and Xu, F. (1992). *Overregularization in Language Acquisition*. Chicago: University of Chicago Press.

Marsden, H. (2004). *Quantifier Scope in Non-Native Japanese: A Comparative Interlanguage Study of Chinese, English and Korean-speaking Learners*. Unpublished PhD dissertation. University of Durham.

May, R. (1977). *The Grammar of Quantification*. PhD dissertation. MIT.

McDonald, J.L. (2000). 'Grammaticality judgments in a second language: Influences of age of acquisition and native language.' *Applied Psycholinguistics* 21:395-423.

McLaughlin, B. (1978). *Second-Language Acquisition in Childhood*. Hillsdale, NJ: Lawrence Erlbaum.

Meisel, J.M., Clahsen, H. and Pienemann, M. (1981). 'On determining developmental stages in natural second language acquisition.' *Studies in Second Language Acquisition* 3:109-135.

Meisel, J.M. (1997). 'The acquisition of the syntax of negation in French and German: Contrasting first and second language development.' *Second Language Research* 13:227-263.

Meisel, J.M. (2004). 'The bilingual child.' In T.K. Bhatia and W.C. Ritchie (eds.). *The Handbook of Bilingualism*. Oxford: Blackwell. pp. 91-113.

Miller, J.F. and Chapman, R.S. (1981). 'The relation between age and mean length of utterance in morphemes.' *Journal of Speech and Hearing Research* 24:154-161.

Miller, K. and Schmitt, C. (2004). 'Wide-scope indefinites in English child language.' In J. Van Kampen and S. Baauw (eds.). *Proceedings of Generative Approaches to Language Acquisition 2003*. Utrecht: LOT Occasional Series. pp. 317-328.

Milsark, G. (1977). 'Toward an explanation of certain peculiarities of the existential construction in English.' *Linguistic Analysis* 3:1-29.

Minifie, F.D., Darley, F.L. and Sherman, D. (1963). 'Temporal reliability of seven language measures.' *Journal of Speech and Hearing Research* 6:139-148.

Miyamoto, E.T. and Takahashi, S. (2002). 'Sources of difficulty in processing scrambling in Japanese.' In M. Nakayama (ed.). *Sentence Processing in East Asian Languages*. Standford: CSLI. pp. 167-188.

Molony, C. (1977). ''Ich *bin* sprechen deutsch aber': The sequence of verb and word order acquisition of an American child learning German.' In C. Molony, H. Zobl and W. Stölting (eds.). *Deutsch im Kontakt met Anderen Sprachen [German in Contact with Other Languages]*. Kronberg: Scriptor Verlag. pp. 274-295.

Montens, F. and Sciarone, A.G. (1991). *Nederlands voor Buitenlanders: De Delftse Methode*. Amsterdam: Boom.

Montrul, S. and Slabakova, R. (2003). 'Competence similarities between native and near-native speakers.' *Studies in Second Language Acquisition* 25:351-398.

Musolino, J. (1998). *Universal Grammar and the Acquisition of Semantic Knowledge: An Experimental Investigation into the Acquisition of Quantifier Negation Interaction in English.*. Unpublished PhD dissertation. University of Maryland.

Musolino, J., Crain, S. and Thornton, R. (2000). 'Navigating negative quantificational space.' *Linguistics* 38:1-32.

Musolino, J. and Lidz, J. (to appear). 'Why children aren't universally successful with quantification.' *Linguistics*.

Neeleman, A. (1994). *Complex Predicates*. PhD dissertation. Utrecht University.

Neeleman, A. and Weerman, F. (1996). Flexible syntax: A theory of case and arguments. Ms. Utrecht University.

Neeleman, A. and Weerman, F. (1997). 'L1 and L2 word order acquisition.' *Language Acquisition* 6:125-170.

Neeleman, A. and Reinhart, T. (1998). 'Scrambling and the PF Interface.' In M. Butt and W. Geuder (eds.). *The Projection of Arguments: Lexical and Compositional Factors*. Stanford: CSLI Publications. pp. 309-353.

Neeleman, A. and Weerman, F. (1999). *Flexible Syntax. A Theory of Case and Arguments*. Dordrecht: Kluwer.

Newport, E.L. (1990). 'Maturational constraints on language learning.' *Cognitive Science* 14:11-28.

Nice, M.M. (1925). 'Length of sentences as a criterion of a child's progress in speech.' *Journal of Educational Psychology* 16:370-379.

Obenauer, H.-G. (1984/1985). 'On the identification of empty categories.' *The Linguistic Review* 4:153-202.

Ortega, L. (2000). *Understanding Complexity: The Measurement of Change in the Syntax of Instructed L2 Spanish Learners*. Unpublished PhD dissertation. University of Hawai'i.

Ortega, L. (2003). 'Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing.' *Applied Linguistics* 24:492-518.

Otsu, Y. (1994). 'Early acquisition of scrambling in Japanese.' In B.D. Schwartz and T. Hoekstra (eds.). *Language Acquisition Studies in Generative Grammar*. Amsterdam: John Benjamins. pp. 253-264.

Ouhalla, J. (1990). 'Sentential negation, relativized minimality and the aspectual state of auxiliaries.' *The Linguistic Review* 7:183-231.

Oyama, S. (1978). 'The sensitive period and comprehension of speech.' *Working Papers on Bilingualism* 16:1-17.

Park, H. (2000). '*When*-questions in second language acquisition.' *Second Language Research* 16:44-76.

Parodi, T. (2000). 'Finiteness and verb placement in second language acquisition.' *Second Language Research* 16:355-381.

Partee, B. (1987). 'Many quantifiers.' In J. Groenendijk, D. de Jongh and M. Stokhof (eds.). *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers. Groningen-Amsterdam Studies in Semantics 8*. Dordrecht: Foris. pp. 115-143.

Patowski, M.S. (1980). 'The sensitive period for the acquisition of syntax in a second language.' *Language Learning* 30:449-472.

Patowski, M.S. (1990). 'Age and accent in a second language: A reply to James Emil Flege.' *Applied Linguistics* 11:73-89.

Penfield, W. and Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton, NL: Princeton University Press.

Penner, Z., Tracy, R. and Weissenborn, J. (2000). 'Where scrambling begins: Triggering object scrambling at the early stage in German and Bernese Swiss German.' In S. Powers and C. Hamann (eds.). *The Acquisition of Scrambling and Cliticization*. Dordrecht: Kluwer. pp. 127-164.

Pesetsky, D. (1995). *Zero Syntax: Experiencers and Cascades*. Cambridge, MA: MIT Press.

Philip, W. (2003). The development of obligatorily specific indefinites in L1 Dutch. Ms. Utrecht University.

Phillips, C. (1995). 'Syntax at age two: Crosslinguistic differences.' *MIT Working Papers in Linguistics* 26:325-382.

Pica, T. and Long, M.H. (1986). 'The linguistic and conversational performance of experienced and inexperienced teachers.' In R.R. Day (ed.). *Talking to Learn: Conversation in Second Language Acquisition*. Rowley, MA: Newbury House. pp. 85-98.

Pienemann, M. (1980). 'The second language acquisition of immigrant children.' In S.W. Felix (ed.). *Second Language Development: Trends and Issues*. Tübingen: Gunter Narr. pp. 41-56.

Pienemann, M. (1981). *Der Zweitspracherwerb ausländischer Arbeiterkinder [The Second Language Acquisition of Foreign Workers' Children]*. Bonn: Bouvier.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1986). 'Productivity and conservatism in language acquisition.' In W. Demopoulos and A. Marros (eds.). *Language Learning and Concept Acquisition. Foundational Issues*. Norwood, NJ: Ablex. pp. 54-82.

Pinker, S. (1989). *Learnability and Cognition. The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.

Poeppel, D. and Wexler, K. (1993). 'The full competence hypothesis of clause structure in early German.' *Language* 69:1-33.

Polišenská, D. (2005). Dutch children's acquisition of verbal and adjectival acquisition. An explorative study. Ms. University of Amsterdam.

Pollock, J.-Y. (1989). 'Verb movement, Universal Grammar and the structure of IP.' *Linguistic Inquiry* 20:635-670.

Poser, W.J. (1992). 'Blocking of phrasal constructions by lexical items.' In I.A. Sag and A. Szabolcsi (eds.). *Lexical Matters*. Stanford, CA: CSLI. pp. 111-130.

Prévost, P. (1997). *Truncation in second language acquisition*. Unpublished PhD dissertation. McGill University.

Prévost, P. and White, L. (2000). 'Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement.' *Second Language Research* 16:103-133.

Prévost, P. (2003). 'Truncation and missing inflection in initial child L2 German.' *Studies in Second Language Acquisition* 25:65-97.

Prince, A. and Smolensky, P. (1997). 'Optimality Theory: From neural networks to Universal Grammar.' *Science* 275:1604-1610.

Pulvermüller, F. and Schumann, J.H. (1994). 'Neurobiological mechanisms of language acquisition.' *Language Learning* 44:681-734.

Punt, L. (1998). *Adjectivale Flexie en Verwerving [Adjectival Inflection and Acquisition]*. Unpublished MA thesis. Utrecht University.

Quist, G. and Stuik, D. (2000). *Teach Yourself Dutch Grammar*. London: Hodder and Stroughton.

Radford, A. (1997). *Syntactic Theory and the Structure of English. A Minimalist Approach*. Cambridge: Cambridge University Press.

Ravem, R. (1978). 'Two Norwegian children's acquisition of English syntax.' In E.M. Hatch (ed.). *Second Language Acquisition: A Book of Readings*. Rowley, MA: Newbury House. pp. 148-154.

Reinhart, T. (1976). *The Syntactic Domain of Anaphora*. PhD dissertation. MIT.

Reinhart, T. (1995). 'Interface Strategies.' *OTS Working Papers in Linguistics*.

Reinhart, T. (1996). 'Interface economy: Focus and markedness.' In M. Bierwisch (ed.). *The Role of Economy Principles in Linguistic Theory*. Berlin: Akademie Verlag. pp. 146-169.

Reinhart, T. (1997). 'Quantifier scope: How labor is divided between QR and choice functions.' *Linguistics and Philosophy* 20:335-397.

Richards, B. (1987). 'Type/Token Rations: What do they really tell us?' *Journal of Child Language* 14:201-209.

Rizzi, L. (1993/1994). 'Some notes on linguistic theory and language development: The case of root infinitives.' *Language Acquisition* 3:371-393.

Romaine, S. (2003). 'Variation.' In C. Doughty and M.H. Long (eds.). *The Handbook of Second Language Acquisition*. Oxford: Blackwell. pp. 409-435.

Rondal, J.A., Ghiotto, M., Bredart, S. and Bachelet, J.-F. (1987). 'Age-relation, reliability and grammatical validity of measures of utterance length.' *Journal of Child Language* 14:433-446.

Ross, J.S. (1967). *Constraints on Variables in Syntax*. PhD dissertation. MIT.

Rounds, P. and Kanagy, R. (1998). 'Acquiring linguistic cues to identify AGENT: Evidence from children learning Japanese as a second language.' *Studies in Second Language Acquisition* 20:509-542.

le Roux, C. and Wilsenach, C. (1999). 'Transfer from L1 to L2: A study of parameter values in the Afrikaans of a child L2 learner.' In C. le Roux, D. Aarons and L. Holtzhausen (eds.). *Fokus op taalverwerving. [Focus on language acquisition]*. Stellenbosch: Stellenbosch Papers in Linguistics. pp. 79-109.

Rullman, H. (1995). 'Geen eenheid. [No unit].' *Tabu: Bulletin voor Taalwetenschap* 25:194-197.

Ruys, E. (1992). *The Scope of Indefinites*.. Unpublished PhD thesis. Utrecht University.

Ruys, E. (2001). 'Dutch scrambling and the strong-weak distinction.' *Journal of Comparative Germanic Linguistics* 4:39-67.

Sandt, v.d.R.A. (1992). 'Presupposition projection as anaphora resolution.' *Journal of Semantics* 9:333-377.

Sauerland, U. (1999). 'Erasability and interpretation.' *Syntax* 2:161-188.

Scarborough, H.S. (1990). 'Index of productive syntax.' *Applied Psycholinguistics* 11.

Scarborough, H.S., Rescorla, L., Tager-Flusberg, H., B., Fowler, A.E. and Sudhalter, V. (1991). 'The relation of utterance length to grammatical complexity in normal and language-disorderd groups.' *Applied Psycholinguistics* 12:23-45.

Schaeffer, J.C. (2000). *The Acquisition of Direct Object Scrambling and Clitic Placement: Syntax and Pragmatics*. Amsterdam: John Benjamins.

Schaeffer, J.C. (2005). On direct object scrambling in typical and impaired child language. Paper presented at University of Southern California, USA [2nd February].

Schetter, W.Z. and van der Cruysse, I. (2002). *Dutch. An Essential Grammar*. London: Routledge.

Schouten, C. (1998). *Zinsinitiële Indefiniete Subjecten in hun Context [Sentence-initial Indefinite Subjects in their Context]*. Unpublished PhD dissertation. Catholic University Nijmegen.

Schreiber, T. and Sprouse, R.A. (1998). 'Knowledge of topicalisation and scrambling in English-German interlanguage.' *McGill Working Papers in Linguistics* 13:132-172.

Schumann, J.H. (1975). 'Affective factors and the problem of age in second language acquisition.' *Language Learning* 25:209-235.

Schwartz, B.D. (1990). 'Un-motivating the motivation for the Fundamental Difference Hypothesis.' In H. Burmeister and P. Rounds (eds.). *Variability in Second Language Acquisition*. Eugene, OR: University of Oregan. pp. 667-684.

Schwartz, B.D. and Tomaselli, A. (1990). 'Some implications from an analysis of German word order.' In W. Abraham, W. Kosmeijer and E. Reuland (eds.). *Issues in Germanic Syntax*. Berlin: Mouton de Gruyter. pp. 252-274.

Schwartz, B.D. (1992). 'Testing between UG-based and problem-solving models of L2A: Developmental sequence data.' *Language Acquisition* 2:1-19.

Schwartz, B.D. and Sprouse, R.A. (1994). 'Word order and nominative Case in nonnative language acquisition: A longitudinal study of L1 Turkish German interlanguage.' In T. Hoekstra and B.D. Schwartz (eds.). *Language Acquisition Studies in Generative Grammar*. Amsterdam: John Benjamins. pp. 317-368.

Schwartz, B.D. and Sprouse, R.A. (1996). 'L2 cognitive states and the Full Transfer/Full Access model.' *Second Language Research* 12:40-72.

Schwartz, B.D. and Sprouse, R.A. (2000). 'When syntactic theories evolve: Consequences for L2 acquisition research.' In J. Archibald (ed.). *Second Language Acquisition and Linguistic Theory*. Oxford: Blackwell. pp. 156-186.

Schwartz, B.D. and Sprouse, R.A. (2002). 'Parallels across L1 acquisition and child L2 acquisition.' In H. Marsden, S. Pourcel and M. Whong-Barr (eds.). *Durham Working Papers in Linguistics 8*. pp. 139-152.

Schwartz, B.D. (2003). 'Child L2 acquisition: Paving the way.' In B. Beachley, A. Brown and F. Conlin (eds.). *Proceedings of the 27th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 26-50.

Schwartz, B.D. (2004). 'Why child L2 acquisition?' In J. Van Kampen and S. Baauw (eds.). *Proceedings of Generative Approaches to Language Acquisition 2003*. Utrecht, The Netherlands: LOT Occasional Series. pp. 47-66.

Scovel, T. (2000). 'A critical review of the critical period research.' *Annual Review of Applied Linguistics* 20:213-223.

Shriner, T.H. (1967). 'A comparison of selected measures with psychological scale values of language values.' *Journal of Speech and Hearing Research* 10:828-835.

Shriner, T.H. and Sherman, D. (1967). 'An equation for assessing language development.' *Journal of Speech and Hearing Research* 10:21-40.

Shriner, T.H. (1969). 'A review of mean length of response as a measure of expressive language development in children.' *Journal of Speech and Hearing Disorders* 10:14-48.

Singleton, D. and Ryan, L. (2004). *Language Acquisition: The Age Factor. Second Edition.* Clevedon: Multilingual Matters.

Slabakova, R. (2003). 'Semantic evidence for functional categories in interlanguage grammars.' *Second Language Research* 19:42-75.

Slavoff, G.R. and Johnson, J.S. (1995). 'The effects of age on the rate of learning a second language.' *Studies in Second Language Acquisition* 17:1-16.

Slobin, D.I. (1993). 'Coding child language data for crosslinguistic analysis.' In M.D. Lampert (ed.). *Talking Data. Transcription and Coding in Discourse Research.* Hillsdale, NJ: Erlbaum. pp. 207-219.

Smolensky, P. (1996). 'On the comprehension/production dilemma in child language.' *Linguistic Inquiry* 27:656-683.

Snow, C.E. and Hoefnagel-Höhle, M. (1982). 'The critical period for language acquisition: Evidence from second language learning.' In S. Krashen, R. Scarcella and M.H. Long (eds.). *Child-Adult Differences in Second Language Acquisition.* Rowley, MA: Newbury. pp. 93-111.

Song, H.-J. and Fisher, C. (2005). 'Who's "she"? Discourse prominence influences preschoolers' comprehension of pronouns.' *Journal of Memory and Language* 52:29-57.

Sorace, A. (2000). 'Syntactic optionality in non-native grammars.' *Second Language Research* 16:93-102.

Sorace, A. (2003). 'Near-nativeness.' In C. Doughty and M.H. Long (eds.). *The Handbook of Second Language Acquisition.* Oxford: Blackwell. pp. 130-151.

Sportiche, D. (1992). Clitic constructions. Ms. University of California.

StatSoft, I. (2004). *Electronic Statistics Textbook.* Tulsa, OK: Statsoft. http://www.statsoft.com/textbook/stathome.html

Stern, H.R. (1984). *Essential Dutch Grammar.* Mineola, NY: Dover Publications.

Su, Y. (2001). *Scope and specificity in child language: A cross-linguistic study on English and Chinese.* Unpublished PhD dissertation. University of Maryland.

Takano, Y. (1998). 'Object shift and scrambling.' *Natural Language and Linguistic Theory* 16:817-889.

Thomas, M. (1994). 'Assessment of L2 proficiency in second language acquisition research.' *Language Learning* 44:307-336.

Thráinsson, H. (2001). 'Object shift and scrambling.' In M. Baltin and C. Collins (eds.). *The Handbook of Contemporary Syntactic Theory.* Oxford: Blackwell. pp. 148-202.

Tran, J. (2005a). 'Verb position and verb form in English-speaking children's L2 acquisition of German.' In A. Brugos, M. Clark-Cotton, R. and S. Ha (eds.). *Proceedings of the 29th Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 592-603.

Tran, J. (2005b). 'Word order and verb inflection in English-speaking children's L2 acquisition of German V2.' *University of Hawai'i Working Papers in Linguistics* 36:1-35.

Truckenbrodt, H. (1995). 'Extraposition from NP and prosodic structure.' *Proceedings of NELS* 25:503-517.

Tyler, L.K. (1983). 'The development of discourse mapping processes: The on-line interpretation of anaphoric expressions.' *Cognition* 13:309-341.

Umeda, M. (2001). *Acquisition of Reflexive Binding in English by Child Speakers of Japanese*. Unpublished MA thesis. University of Durham.

Unsworth, S. (2003a). 'Direct object scrambling in adult and child L2 Dutch, *Proceedings of the Fourth Tokyo Conference on Psycholinguistics*. Tokyo: Hituzi Syobo. pp. 265-289.

Unsworth, S. (2003b). 'Testing Hulk & Müller (2000) on crosslinguistic influence: Root Infinitives in a bilingual German/English child.' *Bilingualism: Language and Cognition* 6:143-158.

Unsworth, S. (2004a). 'On the syntax-semantics interface in Dutch: Adult and child L2 acquisition compared.' *International Review of Applied Linguistics* 42:173-187.

Unsworth, S. (2004b). 'Child L1, child L2 and adult L2 acquisition: Differences and similarities.' In A. Brugos, L. Micciulla and C. Smith, E. (eds.). *Proceedings of the 28th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 633-644.

Unsworth, S. (2005). 'Overcoming the poverty-of-the-stimulus: Scrambled indefinites in English-Dutch interlanguage.' In A. Brugos, M. Clark-Cotton, R. and S. Ha (eds.). *Proceedings of the 29th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 627-638.

Vainikka, A. (1993/1994). 'Case in the development of English syntax.' *Language Acquisition* 3:257-325.

Vainikka, A. and Young-Scholten, M. (1994). 'Direct access to X-bar theory: Evidence from Korean and Turkish adults learning German.' In T. Hoekstra and B.D. Schwartz (eds.). *Language Acquisition Studies in Generative Grammar*. Amsterdam: Benjamins. pp. 265-316.

Vainikka, A. and Young-Scholten, M. (1996a). 'Gradual development of L2 phrase structure.' *Second Language Research* 23:7-39.

Vainikka, A. and Young-Scholten, M. (1996b). 'The early stages in adult L2 syntax: Addiitional evidence from Romance speakers.' *Second Language Research* 12:140-176.

Valian, V. (1991). 'Syntactic subjects in the early speech of American and Italian children.' *Cognition* 40:21-81.

Van Boxtel, S. (2005). *Can the Late Bird Catch the Worm? Ultimate Attainment in L2 Syntax*. PhD dissertation. Radboud University Nijmegen.

van de Craats, I. (2001). 'De derde persoon. Eerste- en tweede-taalverwerving vergeleken [The third person. First and second language acquisition compared].' *Toegepaste Taalwetenschap in Artikelen* 66:23-40.

van der Does, J. and de Hoop, H. (1998). 'Type-shifting and scrambled definites.' *Journal of Semantics* 15:393-416.

Van Geenhoven, V. (1998). *Semantic Incorporation and Indefinite Descriptions: Semantic and Syntactic Aspects of Noun Incorporation in West Greenlandic.* Stanford: CSLI Publications.

van Hout, R. and Vermeer, A. (1988). 'Spontane taaldata en het meten van lexicale rijkdom in tweede-taalverwerving. [Spontaneous speech data and measuring methods of lexical richness in second language acquisition.].' *Toegepaste Taalwetenschap in Artikelen* 32:108-122.

van Kalsbeek, A., Huizinga, M. and Kuiken, F. (1998). *Code Nederlands.* Zutphen: ThiemeMeulenhoff.

Vanden Wyngaerd, G. (1989). 'Object shift as an A-movement rule.' *MIT Working Papers in Linguistics* 11:256-271.

Verhoeven, L. and Vermeer, A. (1989). *Diagnose van Kindertaal. Nederlandse Taalvaardigheid van Autochtone en Allochtone Kinderen. [Child Language Diagnosis. Dutch Language Proficiency of Native and Non-Native Children].* Tilburg: Zwijsen.

Vermeer, A. (1986). *Tempo en Stuktuur van Tweede-taalwewerving bij Turkse en Marokkaanse Kinderen [Tempo and Structure of Second Language Acquisition of Turkish and Moroccan Children].* PhD dissertation. University of Tilburg.

Vermeer, A. (1992). 'Exploring the second language learner lexicon.' In L. Verhoeven and J.H.A.L. De Jong (eds.). *The Construct of Language Proficiency.* Amsterdam: Benjamins. pp.

Vermeer, A. (2000). 'Coming to grips with lexical richness in spontaneous speech data.' *Language Testing* 17:65-83.

Vikner, S. (1994). 'Scandinavian object shift and West Germanic scrambling.' In N. Corver and H. van Riemsdijk (eds.). *Studies on Scrambling.* Berlin: Mouton de Gruyter. pp. 487-517.

Wason, P.C. (1965). 'The contexts of plausible denial.' *Journal of Verbal Learning and Verbal Behavior* 4:7-11.

Webelhuth, G. and den Besten, H. (1987). Remnant topicalisation and the constituent structure of the VP in the Germanic SOV languages. Paper presented at the Tenth GLOW colloquium, Venice, March 30-April 1

Weerman, F. (2002). *Dynamiek in Taal en de Explosie van de Neerlandistiek [Dynamics in Language and the Explosion of Dutch Studies]. Inaugural professorial lecture given on 21st June 2002.* Amsterdam: Universiteit van Amsterdam.

Weerman, F., Bisschop, J. and Punt, L. (2003). L1 and L2 acquisition of Dutch adjectival inflection. Paper presented at Generative Approaches of Language Acquisition 2003, Utrecht, The Netherlands [4th September]

Weijnen, A.A. (1937). *Onderzoek Naar de Dialectgrenzen in Noord-Brabant [Research into the Dialect Boundaries in North-Brabant]* Fijnart.

Weissenborn, J. (1990). 'Functional categories and verb movement: The acquisition of German syntax reconsidered.' In M. Rothweiler (ed.). *Spracherwerb und Grammatik. Linguistische Untersuchungen zum Erwerb von Syntax und Morphologie [Language Acquisition and Grammar. Linguistic Investigations into the Acquisition of Syntax and Morphology].* Opladen: Westdeutscher Verlag. pp. 190-223.

Wexler, K. and Culicover, P.W. (1980). *Formal Principles of Language Acquisition.* Cambridge MA: MIT Press.

Wexler, K. (1998). 'Very early parameter setting and the Unique Checking Constraint: A new explanation for the Optional Infinitive stage.' *Lingua* 106:23-79.

White, L. and Genesee, F. (1996). 'How native is near-native? The issue of ultimate attainment in adult second language acquisition.' *Second Language Research* 12:238-265.

White, L. (1998). 'Second language acquisition and Binding Principle B: Child/adult differences.' *Second Language Research* 14:425-439.

White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Whong-Barr, M. and Schwartz, B.D. (2002). 'Morphological and syntactic transfer in child L2 acquisition of the English dative alternation.' *Studies in Second Language Acquisition* 22:579-616.

Wijnen, F. (1997). 'Temporal reference and eventivity in root infinitives.' *MIT Occasional Papers in Linguistics* 12:1-25.

Wijnen, F., Roeper, T. and van der Meulen, H. (2004). 'Discourse binding: Does it begin with nominal ellipsis?' In J. Van Kampen and S. Baauw (eds.). *Proceedings of Generative Approach Language Acquisition 2003*. Utrecht: LOT. pp. 505-516.

Williams, E. (1997). 'Blocking and anaphora.' *Linguistic Inquiry* 28:577-628.

Wode, H. (1978). 'Developmental sequences in naturalistic L2 acquisition.' In E.M. Hatch (ed.). *Second Language Acquisition: A Book of Readings*. Rowley, MA: Newbury House. pp. 101-117.

Wurmbrand, S. (2000). The structure(s) of particle verbs. Ms.

Wykes, T. (1981). 'Inference and children's comprehension of pronouns.' *Journal of Experimental Child Psychology* 32:264-278.

Yuill, N. and Oakhill, J. (1991). *Children's Problems in Text Comprehension: An Experimental Investigation*. Cambridge: Cambridge University Press.

Zuckerman, S. (2001). *The Acquisition of 'Optional' Movement*. PhD dissertation. University of Groningen.

Zwart, C.J.-W. (1995). 'Word order, intonation and noun phrase interpretation in Dutch.' In V. Samiian and J.C. Schaeffer (eds.). *Proceedings of the Western Conference of Linguistics (WECOL 1994)*. Fresno: California State University. pp. 279-289.

Zwart, C.J.-W. (1997). *Morphosyntax of Verb Movement. A Minimalist Approach to the Syntax of Dutch*. Dordrecht: Kluwer.

APPENDIX A

CHILD L2 LITERATURE

Table A provides an overview of (recent) child L2 studies. It includes: (i) (to the best of my knowledge) most child L2 studies carried out within the generative paradigm within the past 15 years, (ii) (to the best of my knowledge) most studies which explicitly compare child L2 with adult L2 (both generative and non-generative), and (iii) child L2 studies where the L2 is Dutch, as in the present work. It does not include: (i) studies which deal exclusively with phonology or the lexicon; or (ii) child L2 studies which focus (predominantly at least) on the psychological and sociological aspects of linguistic development in L2 children and (iii) studies which deal with L2 reading. Note that some of the studies included here examine child L2 data without (explicitly) discussing child L2 acquisition (e.g. Eubank 1993/4; van de Craats, Corver and van Hout 2000). These are included for the sake of completeness.

The organisation of the table is based on the type of data which each study examines rather than, for example, the theoretical question which it addresses or the languages which it examines. It is divided into four sections and within each section, the authors are presented alphabetically. The first section lists studies which use data from L2 children only. The second details studies looking at data from both L2 and L1 children. The third part deals with studies comparing data from L2 children and L2 adults and the final section lists those which analyse data from all three groups. Many of the studies listed in the first section compare their child L2 data with data from either L1 children and/or L2 adults from the literature. Such comparisons are not subsumed in the table. Studies are placed in the different subsections on the basis of the empirical data which they discuss. (For comparisons of child L2 or adult L2 acquisition with bilingual L1 acquisition, see, for example, Bruhn de Garavito (2002), Hulk and Cornips (to appear) and Perani, Paulesu, Sebastian Galles, Dupoux, Dehaene, Bettinardi, Cappa, Fazio and Mehler (1998).)

The first column lists the study. The second and third column detail the L1 and L2, respectively, by means of a three-letter code (see below). The third and fourth column provide age at first exposure and age at time of testing, respectively. For both of these variables, when no specific age was available, length of exposure (LoE) is used instead. A few studies examine data from children whose age at first exposure is between 3 and 4 years old. Although these children would not be classified as L2 children on the definition adopted here, they are included for the sake of completeness. When a study involves different groups or individuals and data are available for each, these are are given separately using the symbols ❶, ❷, etc. to identify different groups and ①, ② to identify different individuals. When L2ers with different L1s are not differentiated, this is because the author(s) treated these as one group. The fifth column lists the type of data which are examined by means of a three-letter code (see below). All data are cross-sectional unless stated otherwise (with an L for 'longitudinal' in column 5). The final column summarises the linguistic phenomenon investigated in each study. When the relevant information was not discernible from the study in question, this is indicated with "??".

The language codes used in columns 2 and 3 are: AFG Afghan, AFK Afrikaans, ALB Albanian, ARA Arabic, BER Berber, BOS Bosnian, CHI Chinese, DAN Danish, DUT Dutch, ENG English, FRE French, GER German, GRK Greek, HBW Hebrew, HIN Hindi, ICE Icelandic, INM Indonesian/Malay, IRI Irish, ITA Italian, JPN Japanese, KOR Korean, KUD Kurdish, MAR Moroccan Arabic, PER Perisan/Dari, RUS Russian, SIN Sinhalese, SPA Spanish, SRN Surinamese, SWE Swedish, TAR Tarafit, THA Thai, TUR Turkish and VTM Vietnamese. The data codes used in column 5 to give the type of data which are used are: ACT act-out, ADIS auditory discrimination, COMP comprehension task, ELP elicited production, GJT(O) (online) grammaticality judgement task, IMI imitation/sentence repetition, PSEL picture selection, SENC sentence combination, SPON spontaneous speech, STOC story/text comprehension, STOP storytelling, TNS translation, VCB receptive and/or productive vocabulary task and WUG wug-test.

Table A. Overview of child L2 studies

Study	L1	L2	n	Age (exposure)	Age (testing)	Type of data	Linguistic phenomenon investigated
Child L2 only							
Appel (1984)	TUR MAR BER	DUT	❶ 15 ❷ 7 ❸ 8	❶ ca. 6;4-11;8 ❷ ca. 6;5-11;10 ❸ ca. 6;5-11;10	❶ LoE: ca. 0;8-2;8 ❷ LoE: ca. 0;8-2;8 ❸ LoE: ca. 0;8-2;8	SPON ELP (L) [a]	Verb placement (amongst other things)
Armon-Lotem (1998)	HBW	ENG	2	① 6;5 ② 4;8	① 6;8-7;2 ② 4;11-5;5	SPON (L)	RIs
Dulay & Burt (1973)	SPA	ENG	151	LoE: <1 yr -??	5-8 yrs	elp	Grammatical morphemes
Dulay & Burt (1974)	SPA CHI	ENG	❶ 60 ❷ 55	❶ ?? ❷ ??	❶ 6-8 yrs ❷ 6-8 yrs	ELP	Grammatical morphemes
Flanigan (1995)	CHI INM KOR ICE ARA SPA SIN HBW	ENG	23	LoE: 0;4-1;9	6;6-14 yrs	COMP SENC	Anaphora (reflexives, pronouns) Relativisation
Gavruseva (1998)	RUS	ENG	2	① 5;2 ② 4;9	① 5;11 ② 6;5	ELP	Constraints on extraction out of possessive NPs
Gavruseva (2000)	RUS	ENG	1	8 yrs	LoE: 0;2-0;8	SPON (L)	RIs
Gavruseva (2002)	RUS	ENG	1	8 yrs	LoE: 0;2-0;8	SPON (L)	Tense and aspect
Gavruseva & Lardière (1996)	RUS	ENG	1	8 yrs	LoE: 0;2-0;8	SPON (L)	Functional categories (CP)
Grondin & White (1996)	ENG	FRE	2	① 4;9 ② 4;5 [b]	① 5;10-8;1 ② 5;6-7;9	SPON (L)	Functional categories (DP, IP, CP)
Haberzettl (1999)	TUR RUS	GER	❶ 2 ❷ 1	6 yrs 8 yrs	LoE: 0;6-1;4 [c] LoE: 0;3-1;3	SPON ELP	Verb placement

[a] Subjects were tested three times within a period of two years.
[b] For approximately the first year of exposure, these children were in a bilingual English-French nursery programme, but they produced very few spontaneous French utterances during this period.
[c] The upper figure is based on the sample numbers (which measure number of months of contact) given in the paper.

Table A (cont'd). Overview of child L2 studies

Study	L1	L2	n	Age (exposure)	Age (testing)	Type of data	Linguistic phenomenon investigated
Child L2 only (cont'd)							
Haznedar (1997)	TUR	ENG	1	4;2	4;3-5;9	SPON (L)	VP and negation
Haznedar (2001)	TUR	ENG	1	4;2	4;3-5;9	SPON (L)	Functional categories (IP)
Haznedar (2003)	TUR	ENG	1	4;2	4;3-5;9	SPON (L)	Functional categories (CP)
Haznedar & Schwartz (1997)	TUR	ENG	1	4;2	4;3-5;9	SPON (L)	RIs
Henry & Tangney (1996)	ENG	IRI	??	3 yrs	??	??	RIs
Ionin & Wexler (2002)	RUS	ENG	20	0;3-13;8 [d]	3;9-13;10	ELP GJT	Tense/agreement morphology
Kroffke & Rothweiler (to appear)	TUR	GER	❶ 2 ❷ 5	❶ 6 yrs ❷ 2;8-4;4	❶ LoE: 0;6 - … [e] ❷ LoE: 0;3-0;9 -…	SPON (L)	Sentential negation Verb placement Subject-verb agreement
Lakshmanan (2000)	CHI	ENG	3	4;8-5;9	5;0-6;1	SPON	Inflectional morphology and Case-checking
Park (2000)	KOR	ENG	6	3;10-8;10	4;7-9;7 [f]	SPON ELP (L)	Wh-questions (formed with *when*)
Prévost (2003)	ENG	GER	1	3;2	3;2-3;7	SPON (L)	RIs
Rounds & Kanagy (1998)	ENG	JPN	89	5-6 yrs	5-12 yrs	PSEL	Word order and Case-marking
Roux & Wilsenach (1999)	ENG	AFK	1	4;6	4;9 – 5;4	SPON (?) (L)	Verb placement

[d] Some of these subjects had had previous exposure to English in Russia.
[e] It is not clear how long data collection continued for.
[f] This is the range of ages for all subjects at the first recording. Subjects were recorded over a period of three years.

Table A (cont'd). Overview of child L2 studies

Study	L1	L2	n	Age (exposure)	Age (testing)	Type of data	Linguistic phenomenon investigated
Child L2 only (cont'd)							
Schwartz & Sprouse (2002)	ENG ITA RUS	FRE GER ENG	2 2 1	① ca. 5;2 ② ca. 5;6 ① 8 yrs ② 8yrs 8 yrs	① ca. 5;2-7;8 ② ca. 5;6-8;0 ① LoE: 1wk-1;2.15 ② LoE: 3wks-1;6 LoE: 0;2-0;8	SPON (L)	RIs
Tran (2005a; 2005b)	ENG	GER	14	4-5 yrs	8;11-14;0	ELP	V2 and RIs
Vermeer (1986)	TUR MAR	DUT DUT	❶ 16 ❷ 16	❶ LoE: 5;7.15 [g] ❷ LoE: 4;3.15	❶ 6;11.15-9;1.15 ❷ 6;9.15-8;11.15	SPON ELP VCB IMI (L)	General proficiency
White (1996)	ENG	FRE	2	① 5;4 ② 5;10	① 5;9-7;9 ② 5;10-8;1	SPON	Clitics
Child L2 cf. child L1							
Aarssen (1996)	TUR DUT TUR	DUT --- ---	❶ 50 ❷ 50 ❸ 75	❶ 4 yrs ?? [h] ❷ --- ❸ ---	❶ 4-10 yrs ❷ 4-10 yrs ❸ 5-9 yrs	ELP PSEL COMP	Anaphors Relative clauses Narrative discourse (topics and temporality)
Bos (1997)	MAR DUT MAR	DUT --- ---	❶ 50 ❷ 50 ❸ 75	❶ LoE: >2 yrs ❷ --- ❸ ---	❶ 4-10 yrs ❷ 4-10 yrs ❸ 5-9 yrs	ELP PSEL COMP	Anaphors Relative clauses Narrative discourse (topics and temporality)
Håkansson (2001)	alb and /or bos ARA SWE SWE (SLI)	SWE --- ---	❶ 10 ❷ 10 ❸ 10	❶ ≥ 3 yrs ❷ --- ❸ ---	❶ 3;6-6;0 [i] ❷ 3;1-3;7 ❸ 4;10-6;3	ELP (L)	Tense morphology and V2

[g] The figures for the child subjects are averages calculated using the averages given in Vermeer (1986).
[h] These children mostly used Turkish at home and started learning Dutch when they started primary school at four years. From the description in Aarssen (1996), however, it is not clear whether they were exposed to Dutch prior to this.
[i] Subjects had had four months' exposure at the first recording and they were tested a second time when approximately six months older.

Table A (cont'd). Overview of child L2 studies

Study	L1	L2	n	Age (exposure)	Age (testing)	Type of data	Linguistic phenomenon investigated
Child L2 cf. child L1 (cont'd)							
Jin (2003)	KOR ENG	ENG ---	❶ 30 ❷ 10 ❸ 10	❶ LoE: 0;5-1;8 ❷ --- ❸	❶ 5-8 yrs ❷ 4 yrs ❸ 6 yrs	IMI	DP
Lalleman (1986)	TUR DUT	DUT ---	❶ 20 ❷ 20	❶ ca.4-5 yrs [j] ❷ ---	❶ 5;11-7;2 ❷ 6;0-7;8	ELP IMI ACT	General proficiency
Paradis, LeCorre & Genessee (1998)	ENG FRE	FRE ---	❶ 15 ❷ 5	❶ 5 yrs ❷ ---	❶ 6-7 yrs ❷ 6-7 yrs	ELP (L) [k]	Tense/ agreement morphology
Shin & Milroy (1999)	KOR ENG	ENG ---	❶ 12 ❷ 2	❶ 5 yrs (?) [l] ❷ ---	❶ 6;6-7;4 ❷ classmates	ELP	Grammatical morphemes
Verhoeven & Vermeer (1984) [m]	TUR SRN DUT	DUT DUT ---	❶ 92 ❷ 33 ❸ 83	❶ 4 yrs ❷ 4 yrs ❸ ---	❶ 4-8 yrs ❷ 4-8 yrs ❸ 4-8 yrs	ADIS IMI VCB ELP PSEL STOC	General proficiency
Whong-Barr & Schwartz (2002)	JPN KOR ENG	ENG ENG ---	❶ 5 ❷ 5 ❸ 6	❶ 4;2-7;8 ❷ 3;7-8;1 ❸ ---	❶ 7;3-8;11 ❷ 6;6-10;2 ❸ 6;11-10;10	GJT	Dative alternation

[j] The author does not provide the age at first exposure for each subject. The figure given here is calculated by using the data for age at time of testing and number of years of nursery school attendance (Lalleman 1986:6, Table 1).
[k] Three recordings were made over a period of three years.
[l] No specific information is provided except that all children (except one) attended kindergarten in the US.
[m] This study constitutes part of the development of the *Toets Allochtone Kinderen (TAK)* 'Immigrant children test' (see Verhoeven and Vermeer 1989 and Chapter 4 for more details).
[n] These children were born in The Netherlands and will therefore probably have had some exposure to Dutch before the age of four (E. Blom, personal communication, 23 June, 2005).

Table A (cont'd). Overview of child L2 studies

Study	L1	L2	n	Age (exposure)	Age (testing)	Type of data	Linguistic phenomenon investigated
Child L2 cf. adult L2							
Blom & Polišenská (to appear)	MAR and /or TAR TUR	DUT DUT	❶ 17 ❷ 6 ❸ 10 ❹ 5	❶ ca.4 yrs [n] ❷ > 15 yrs ❸ ca.4 yrs ❹ > 15 yrs	❶ 5-8 yrs ❷ 22-44 yrs ❸ 5-8yrs ❹ 22-44 yrs	ELP	Finiteness
Cancino, Rosansky & Schumann (1978)	SPA	eng	❶ 2 ❷ 2 ❸ 2	① 4 yrs ② 5 yrs ① 10 yrs ② 12 yrs ① 25 yrs ② 33 yrs	① LoE: 0;1 [p] ② LoE: 0;4 ① LoE: 0;1[o] ② LoE: 0;1 ① LoE: 0;3 [q] ② LoE: 0;4	SPON IMI ELP (L)	Negation Interrogatives
Dimroth (submitted)	RUS	GER	2	① 8;7 ② 14;2	① LoE: 3 weeks ② LoE: 3 weeks	SPON (L)	Finiteness and negation
Eubank (1993/4)	FRE FRE	ENG ENG	❶ 1 ❷ 2	❶ ?? ① ?? ② ??	❶ 4;6-5;6 ① 9;2-9;11 ② 11;1-11;10	SPON (L)	Verb raising and negation
Hilles (1991)	SPA	ENG	❶ 2 ❷ 2 [r] ❸ 2	① 4 yrs ② 5 yrs ① 10 yrs ② 12 yrs ① 25 yrs ② 33 yrs	① LoE : 0;1[s] ② LoE: 0;4 ① LoE: 0;1 ② LoE: 0;1 ① LoE: 0;3 [t] ② LoE: 0;4	SPON (L)	Pronominal subjects and inflection
Kessler & Idar (1979)	VTM	ENG	2	① 4;0 ② late 20s	① 4;6 ② LoE: 1;2	SPON (L)	Grammatical morphemes

[o] Both these subjects had some (limited) previous exposure to English in the form a tutor in their home country.
[p] This child (Marta) had previously also been exposed to some (limited) English in nursery school in Puerto Rico and at a summer camp.
[q] This subject (Dolores) had studied English at high school and university in Peru prior to arrival in the United States.
[r] Both subjects had some (limited) previous exposure to English in the form a tutor in their home country.
[s] This child (Marta) had previously also been exposed to some (limited) Eng in nursery school in Puerto Rico and at a summer camp.
[t] This subject (Dolores) had studied Eng at high school and university in Peru prior to arrival in the United States.

Table A (cont'd). Overview of child L2 studies

Study	L1	L2	n	Age (exposure)	Age (testing)	Type of data	Linguistic phenomenon investigated
Child L2 cf. adult L2 (cont'd)							
Van de Craats, Corver & van Hout (2000)	TUR MAR	DUT DUT	❶ 16 ❷ 2 ❸ 16 ❹ 2	❶ LoE: 5;7.15 [u] ① 19 yrs ② 17 yrs ❸ LoE: 4;3.15 ① 24 yrs ② 19 yrs	❶ 6;11.15-9;1.15 ① LoE: 0;9-2;3 ② LoE: 0;11-2;5 ❸ 6;9.15-8;11.15 ① LoE: 1;2-2;8 ② LoE: 0;8-2;2	SPON ELP (L)	Possessive NPs
Child L2 cf. adult L2 cf. child L1							
Gilkerson (2004)	SPA ENG	ENG ---	❶ 32 ❷ 33 ❸ 8	❶ ca. 5 yrs ❷ > 18 yrs ❸ ---	❶ 4;4-8;7 ❷ 21-45 yrs ❸ 4;9-6;11	ELP	Particle verbs
McDonald (2000)	SPA VTM ENG	ENG ENG ---	❶ 14 ❷ 14 ❸ 14 ❹ 10 ❺ 14	❶ 3-5 yrs ❷ 4-19 yrs [v] ❸ 3-5 yrs ❹ 6-10 yrs ❺ ---	❶ 18-26 yrs ❷ 21-25 yrs ❸ 18-23 yrs ❹ 18-24 yrs ❺ ??	GJTO	Various morpho-syntactic rules
Snow & Hoefnagel-Höhle (1982)	ENG DUT	DUT ---	❶ 10 ❷ 14 ❸ 19 ❹ 17 ❺ 21 ❻ 8 ❼ 8	LoE: < 0;1.5 for beginners; ?? for advanced group ❻ --- ❼ ---	❶ 3-5 yrs ❷ 6-7 yrs ❸ 8-10 yrs ❹ 12-15 yrs ❺ adults ❻ 6-7 yrs ❼ 12-15 yrs	IMI ADIS WUG IMI TNS GJT VCB STOC STOP (L)	Various aspects of phonology, morphology and syntax
Tsimpli & Mastropa-vlou (to appear)	TUR TUR RUS GRK (SLI)	GRK GRK ---	❶ 10 ❷ 6 ❸ 6	❶ 6 yrs [w] ❷ 18-35 yrs ❸ ---	❶ 8-12 yrs ❷ LoE: > 8 yrs ❸ 4;0-6;2	SPON ELP	Determiners Clitics

[u] The figures for the child subjects are averages which are calculated using the averages given in Vermeer (1986), the corpora which Van de Craats *et al* (2000) use.
[v] All subjects, except two, were exposed to English between 3 and 5 years. The figures given here are taken from the 'self-given age of acquisition' column in McDonald's (2000) Tables 1 and 4.
[w] The exact age of first exposure for these L2 children is not entirely clear. The authors state that at the age of 6 years, when schooling began, they had a minimal knowledge of Greek, but there were nevertheless large enough differences between children to be assigned to different school classes.

Table A (cont'd). Overview of child L2 studies

Study	L1	L2	n	Age (exposure)	Age (testing)	Type of data	Linguistic phenomenon investigated
Child L2 cf. adult L2 cf. child L1 (cont'd)							
Van de Craats (2001)	TUR MAR DUT	DUT DUT ---	❶ 46 ❷ 4 ❸ 41 ❹ 4 ❺ 61	❶ 4 yrs-8;4 ❷ 17-19 yrs ❸ LoE: >2 yrs ❹ 18-24 yrs ❺ ---	❶ 4-10 yrs ❷ LoE: 0;9-2;6 ❸ 4-10 yrs ❹ LoE: 0;7-2;8 ❺ 1;6-9 yrs	ELP SPONT (L) [x]	Third person possessive constructions
Weerman (2002) Weerman, Bisschop & Punt (2003)	BER ITA TUR CHI FRE KUD HIN BER THA PER DAN CHI MAR TUR DUT	DUT ---	❶ 6 ❷ 14 ❸ 20	❶ 4 yrs ❷ 14-59 yrs [y] ❸ ---	❶ 16-18 yrs ❷ 16-61 yrs ❸ 3-7 yrs	ELP	Adjectival inflection

[x] Some of the data reported on are cross-sectional.

[y] These figures were calculated by subtracting the length of exposure from the age at time of testing. It should be noted that in the original data – taken from Punt (1998) – one of the subjects who was classified as an adult L2er appeared to be a child L2er and hence is classified as such here. This subject was 16 years old when tested and had had 12 years of instruction. I assume therefore that his age at first exposure is 4 years.

References for Table A

Aarssen, J. (1996). *Relating events in two languages. Acquisition of cohesive devices by Turkish-Dutch bilingual children at school age*. PhD dissertation. University of Tilburg.

Appel, R. (1984). *Immigrant children learning Dutch. Sociolinguistic and psycholinguistic aspects of second-language acquisition*. Dordrecht: Foris.

Armon-Lotem, S. (1998). 'Root Infinitives in child second language acquisition.' *McGill Working Papers in Linguistics* 13:11-17.

Blom, E. and Polišenská, D. (to appear). 'Verbal inflection and verb placement in first and second language acquisition.' In M. Vliegen (ed.). *Proceedings of the 29th Linguistics Colloquium*. Amsterdam: Peter Lang.

Bos, P. (1997). *Development of bilingualism. A study of school-age Moroccan children in the Netherlands*. PhD dissertation. University of Tilburg.

Bruhn de Garavito, J. (2002). 'Verb raising in Spanish: A comparison of early and late bilinguals.' In B. Skarabela, S.A. Fish and H.-J.A. Do (eds.). *Proceedings of the 26th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 84-94.

Cancino, H., Rosansky, E.J. and Schumann, J.H. (1978). 'The acquisition of English negatives and interrogatives by native Spanish speakers.' In E.M. Hatch (ed.). *Second Language Acquisition. A Book of Readings*. Rowley, MA: Newbury House. pp. 207-230.

Dimroth, C. (submitted). 'Age effects on the process of L2 acquisiton? Evidence from the acquisition of negation and finiteness in L2 German.' *Second Language Research*.

Dulay, H. and Burt, M. (1974). 'Natural sequences in child second language acquisition.' *Language Learning* 24:37-53.

Dulay, H.C. and Burt, M.K. (1973). 'Should we teach children syntax?' *Language Learning* 23:244-258.

Eubank, L. (1993/4). 'On the transfer of parametric values in L2 development.' *Language Acquisition* 3:183-208.

Flanigan, B.O. (1995). 'Anaphora and relativization in child second language acquisition.' *Studies in Second Language Acquisition* 17:331-351.

Gavruseva, E. and Lardiere, D. (1996). 'The emergence of extended phrase structure.' In A. Stringfellow, D. Cahana-Amitay, E. Hughes and A. Zukowski (eds.). *Proceedings of the 20th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 225-236.

Gavruseva, E. (1998). 'Left-branch violations in child L2 English.' In A. Greenhill, M. Hughes, H. Littlefield and H. Walsh (eds.). *Proceedings of the 22nd Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 235-245.

Gavruseva, E. (2000). 'Aspect parameter in the guise of Optional Infinitives in child L2 English.' In S.C. Howell, S.A. Fish and T. Keith-Lucas (eds.). *Proceedings of the 24th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 319-330.

Gavruseva, E. (2002). 'Aspect, eventivity and finiteness in early child grammar.' In J. Costa and J. Freitas (eds.). *Proceedings of Generative Approaches to Language Acquisition 2001*. Lisbon: Associação Portuguesa de Linguística. pp. 68-75.

Gilkerson, J. (2004). *Acquiring English particle verbs: Age and transfer effects in L2 acquisition*. Paper presented at the inaugural Generative Approaches to Language Acquisition North America (GALANA), University of Hawai'i, Honolulu, USA.

Grondin, N. and White, L. (1996). 'Functional categories in child L2 acquisition of French.' *Language Acquisition* 5:1-34.

Haberzettl, S. (1999). '*Katze Maus essen* vs. *Katze essen Maus*: Die L1 als Königs- oder Holzwegbereiter zur L2? Zum Einfluss des L1-Wissens im Erwerb der deutschen Verbstellung durch türkische und russische Kinder.' In H.O. Spillmann and I.

Warnke (eds.). *Internationale Tendenzen der Syntaktik, Semantik und Pragmatik [International Trends in Syntax, Semantics and Pragmatics]*. Frankfurt: Peter Lang. pp. 157-165.

Håkansson, G. (2001). 'Tense morphology and verb-second in Swedish L1 children, L2 children and children with SLI.' *Bilingualism: Language and Cognition* 4:85-99.

Haznedar, B. (1997). *Child L2 acquisition of English: A longitudinal case study of a Turkish-speaking child*. Unpublished PhD dissertation. University of Durham.

Haznedar, B. and Schwartz, B.D. (1997). 'Are there Optional Infinitives in child L2 acquisition?' In E. Hughes, M. Hughes and A. Greenhill (eds.). *Proceedings of the 21st Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 257-268.

Haznedar, B. (2001). 'The acquisition of the IP system in child L2 English.' *Studies in Second Language Acquisition* 23:1-39.

Haznedar, B. (2003). 'The status of functional categories in child second language acquisition: Evidence from the acquisition of CP.' *Second Language Research* 19:1-41.

Henry, A. and Tangney, D. (1996). 'Functional categories in child L2 acquisition of Irish.' In A. Stringfellow, D. Cahana-Amitay and A. Zukowski (eds.). *Proceedings of the 20th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 318-127.

Hilles, S. (1991). 'Access to Universal Grammar in second language acquisition.' In L. Eubank (ed.). *Point/counterpoint: Universal Grammar in the second language*. Amsterdam: John Benjamins. pp. 305-338.

Hulk, A. and Cornips, L. (to appear). 'The acquisition of definite determiners in child L2 Dutch: Problems with neuter gender nouns.' In S. Unsworth, T. Parodi, A. Sorace and M. Young-Scholten (eds.). *Paths of Development in L1 and L2 acquisition*. Amsterdam: John Benjamins.

Ionin, T. and Wexler, K. (2002). 'Why is 'is' easier than '-s?: Acquisition of tense/agreement morphology by child second language learners of English.' *Second Language Research* 18:95-136.

Jin, J. (2003). 'Acquisition of L2 English DP by Korean children.' *Reading Working Papers in Linguistics* 7:77-102.

Kessler, C. and Idar, I. (1979). 'Acquisition of English by a Vietnamese mother and child.' *Working Papers in Bilingualism* 18:65-79.

Kroffke, S. and Rothweiler, M. (to appear). 'Variation im frühen Zweitspracherwerb des Deutschen durch Kinder mit türkischer Erstsprache.' In M. Vliegen (ed.). *Proceedings of the 39th Linguistics Colloquium*. Amsterdam: Peter Lang. pp.

Lakshmanan, U. (2000). 'Clause structure in child second language grammars.' *University of Pittsburgh Wokring Papers in Linguistics* 4:15-39.

Lalleman, J. (1986). *Dutch Language Proficiency of Turkish Children Born in The Netherlands*. Dordrecht: Foris.

McDonald, J.L. (2000). 'Grammaticality judgments in a second language: Influences of age of acquisition and native language.' *Applied Psycholinguistics* 21:395-423.

Paradis, J., Le Corre, M. and Genesee, F. (1998). 'The emergence of tense and agreement in child L2 French.' *Second Language Research* 14:227-256.

Park, H. (2000). '*When*-questions in second language acquisition.' *Second Language Research* 16:44-76.

Perani, D., Paulesu, E., Sebastian Galles, N., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S.F., Fazio, F. and Mehler, J. (1998). 'The bilingual brain. Proficiency and age of acquisition of the second language.' *Brain* 121:1841-1852.

Prévost, P. (2003). 'Truncation and missing inflection in initial child L2 German.' *Studies in Second Language Acquisition* 25:65-97.

Punt, L. (1998). *Adjectivale Flexie en Verwerving [Adjectival Inflection and Acquisition]*. Unpublished MA thesis. Utrecht University.

Rounds, P. and Kanagy, R. (1998). 'Acquiring linguistic cues to identify AGENT: Evidence from children learning Japanese as a second language.' *Studies in Second Language Acquisition* 20:509-542.

le Roux, C. and Wilsenach, C. (1999). 'Transfer from L1 to L2: A study of parameter values in the Afrikaans of a child L2 learner.' In C. le Roux, D. Aarons and L. Holtzhausen (eds.). *Fokus op taalverwerving. [Focus on language acquisition]*. Stellenbosch: Stellenbosch Papers in Linguistics. pp. 79-109.

Schwartz, B.D. and Sprouse, R.A. (2002). 'Parallels across L1 acquisition and child L2 acquisition.' In H. Marsden, S. Pourcel and M. Whong-Barr (eds.). *Durham Working Papers in Linguistics 8*. pp. 139-152.

Shin, S.J. and Milroy, L. (1999). 'Bilingual language acquisition by Korean school children in New York City.' *Bilingualism: Language and Cognition* 2:147-167.

Snow, C.E. and Hoefnagel-Höhle, M. (1982). 'The critical period for language acquisition: Evidence from second language learning.' In S. Krashen, R. Scarcella and M.H. Long (eds.). *Child-Adult Differences in Second Language Acquisition.* Rowley, MA: Newbury. pp. 93-111.

Tran, J. (2005a). 'Verb position and verb form in English-speaking children's L2 acquisition of German.' In A. Brugos, M. Clark-Cotton, R. and S. Ha (eds.). *Proceedings of the 29th Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press. pp. 592-603.

Tran, J. (2005b). 'Word order and verb inflection in English-speaking children's L2 acquisition of German V2.' *University of Hawai'i Working Papers in Linguistics* 36:1-35.

Tsimpli, I. and Mastropavlou, M. (to appear). 'Feature-interpretability in child and adult L2 acquisition and SLI.' In J. Liceras, H. Zobl and H. Goodluck (eds.). *Features, Minimalism and Language Acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates.

van de Craats, I., Corver, N. and van Hout, R. (2000). 'Conservation of grammatical knowledge: On the acquisition of possessive noun phrases by Turkish and Morroccan learners of Dutch.' *Linguistics* 38:221-314.

van de Craats, I. (2001). 'De derde persoon. Eerste- en tweede-taalverwerving vergeleken [The third person. First and second language acquisition compared].' *Toegepaste Taalwetenschap in Artikelen* 66:23-40.

Verhoeven, L. and Vermeer, A. (1984). 'Ethnic group differences in children's oral proficiency of Dutch.' In G. Extra and T. Vallen (eds.). *Ethnic minorities and Dutch as a Second Language.* Dordrecht: Foris. pp. 105-131.

Verhoeven, L. and Vermeer, A. (1989). *Diagnose van Kindertaal. Nederlandse Taalvaardigheid van Autochtone en Allochtone Kinderen. [Child Language Diagnosis. Dutch Language Proficiency of Native and Non-Native Children]*. Tilburg: Zwijsen.

Vermeer, A. (1986). *Tempo en Stuktuur van Tweede-taalvewerving bij Turkse en Marokkaanse Kinderen [Tempo and Structure of Second Language Acquisition of Turkish and Moroccan Children]*. PhD dissertation. University of Tilburg.

Weerman, F. (2002). *Dynamiek in Taal en de Explosie van de Neerlandistiek [Dynamics in Language and the Explosion of Dutch Studies]. Inaugural professorial lecture given on 21st June 2002.* Amsterdam: Universiteit van Amsterdam.

Weerman, F., Bisschop, J. and Punt, L. (2003). L1 and L2 acquisition of Dutch adjectival inflection. Paper presented at Generative Approaches of Language Acquisition 2003, Utrecht, The Netherlands [4th September]

White, L. (1996). 'Clitics in L2 French.' In H. Clahsen (ed.). *Generative Perspectives on L2 Acquisition.* Amsterdam: John Benjamins. pp. 335-368.

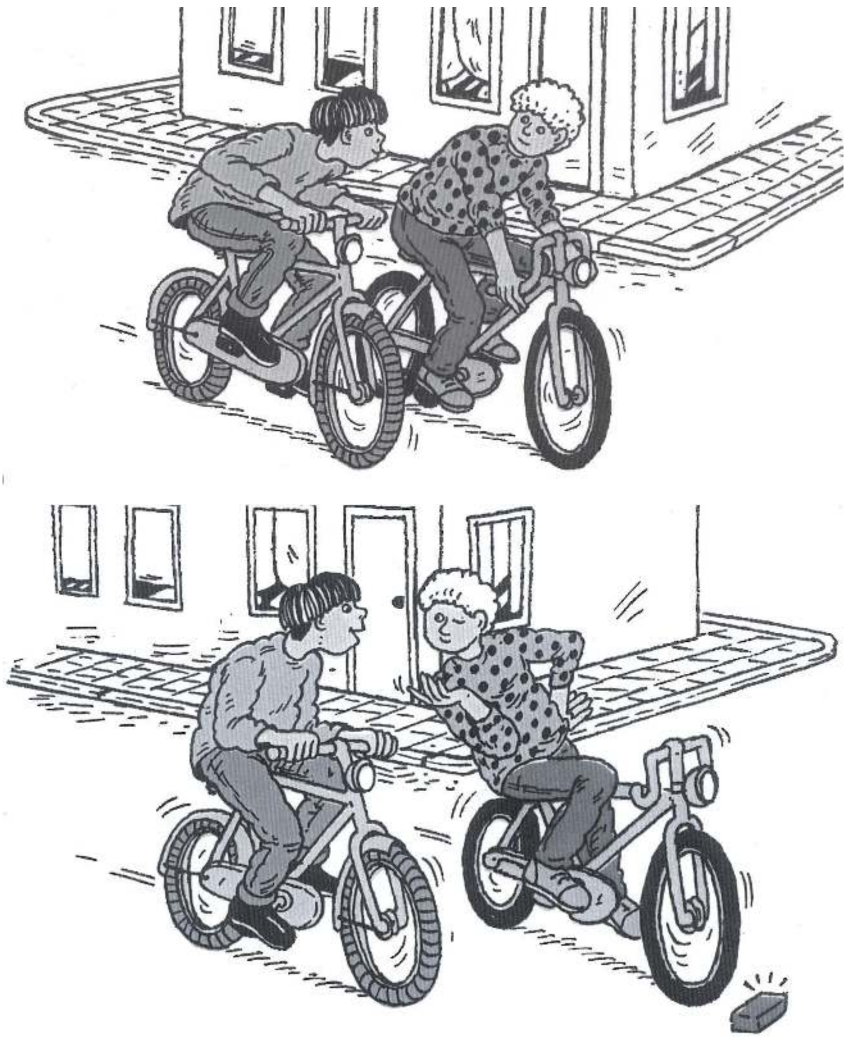
Whong-Barr, M. and Schwartz, B.D. (2002). 'Morphological and syntactic transfer in child L2 acquisition of the English dative alternation.' *Studies in Second Language Acquisition* 22:579-616.

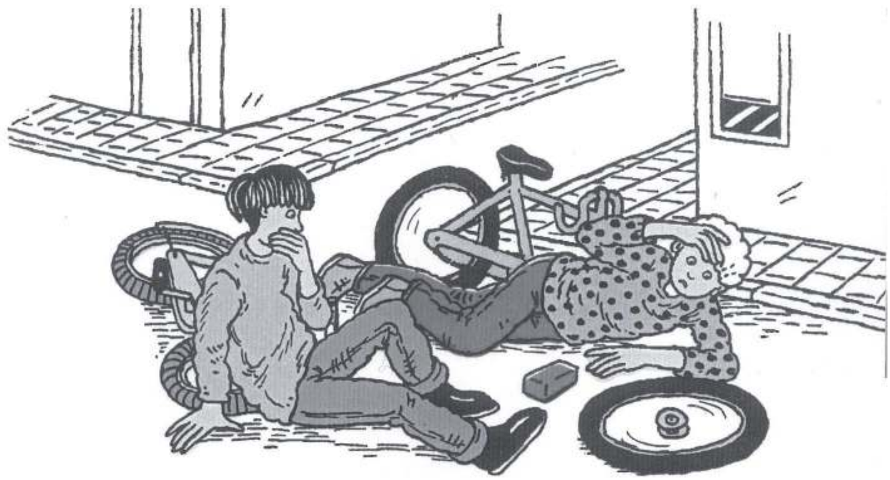
APPENDIX B

PROFICIENCY MEASURE (CHAPTER 4)

Example picture story for picture description task

Source: van der Zee, J., Nales, B. and Smit, J. *Praatboek: Naar de Dierentuin [Talkbook: To the Zoo]*. Doetinchem: Edudesk.





Example transcript of coded proficiency data

@Begin
@Languages: Dutch, English
@Participants: SUB Subject, WIL Wil Student, SHA Sharon Investigator
@Age of SUB: 10;3.0
@Birth of SUB: 07-MAR-1994
@Date: 07-JUN-2004

@Situation: boy and ball
*SUB: de jongen heeft de bal aan de fast [?] +//. [+ inc]
*SUB: nee. [+ excl]
*SUB: de jongen heeft de bal gegooid en gaat weg. [+ EF]
*SUB: en de hond ziet de bal. [+ EF]
*SUB: en de jongen huilt. [+ EF]
*SUB: zo de hond ga naar de bal.
*SUB: het [//] de bal gaat op de straat.
%com: target: "rolt/stuitert de straat op"
*SUB: and hij breng de bal terug naar de jongen.
*SUB: en de jongen <gooit de bal weg> [/] gooit de bal weg. [+ EF]

@Situation: Paddington growing flowers
*SUB: Paddington+Beer brengt bloemen. [+ EF]
*SUB: hij doet water over de bloemen. [+ EF]
*SUB: and <een man> [/] een [/] een man kijkt over de hek. [+ EF]
*SUB: <en ziet> [//] en hij ziet de beer. [+ EF]
*SUB: hij doet water op de bloemen nog een keer.
%com: target: "...nog een keer water op de bloemen"
*SUB: en hij geeft één bloem <naar de man> [/] naar de man.
%com: target: "aan" instead of "naar"

@Situation: Grover and puppet show
*SUB: de blauwe ding hij doet sokken <in zijn> [/] <in zijn> [//] +//. [+ inc]
*SHA: la(de).
*WIL: in de la(de) bedoel je of in z(ij)n hand ?
*SUB: dit. [+ EF] [+ ell]
%act: points to chest of drawers
*WIL: de la(de).
*SUB: hij doet sokken +/.
*WIL: haalt sokken uit de la(de) of uit de kast.
*SUB: +, uit de kast.
%com: target: "haalt" instead of "doet"
*SUB: <hij maakt> [//] ze maakt [//] ze pakt schaar lijm en +/.
%com: target: "een/de schaar"
*WIL: ++ wol.
*SUB: +, wol.
*SUB: wat is dit? [+ form]
%act: points to wool
%com: order of utterances changed so that CLAN counts utterance ending with +/. and utterance beginning with +, as one
*SUB: <en ze maakt> [//] en ze doet iets met de sokken. [+ EF]

Example transcript of coded proficiency data (cont'd)

*SUB: en dan maak [?] ze poppen. [+ excl]
%com: "maak" could be "maakt"
*SUB: en ze doet een pop [//] +/.
*WIL: ++ poppenkast.
*SUB: +, poppenkast. [+ EF]

@Situation: Ernie being painted
*SUB: die jongen doet <één # pet> [/] +/.
*SUB: +, een pet op de meisjes hoofd. [+ EF]
*SUB: nee. [+ excl]
*SUB: dat is niet een pet. [+ scr]
*WIL: het wel een soort pet.
%com: order of utterances changed so that CLAN counts utterance ending with +/. and utterance beginning with +, as one
*SUB: dan rijden hij zijn vriend.
*SUB: en dan schrijft die [/] <de meisje> [/] de meisje +... [+ inc]
*WIL: wat heeft ze gemaakt?
*SUB: een +... [+ inc]
*WIL: een schilderij?
*SUB: ja. [+ excl]
*SUB: een schilderij van de jongen en zijn paard. [+ EF] [+ ell]

@Situation: Paddington baking cookies
*SUB: Paddington+Bear <hij helpt zijn> [//] hij [?] helpt een vrouw xx cookies te maken. [+ excl]
*SUB: dan maakt hij <in de> [//] ## een hart. [+ EF]
*SUB: de vrouw # ze pakt koekjes <uit de> [/] <uit de> [/] <uit de> [//] +/.
*WIL: ++ uit de oven.
*SUB: +, uit het oven. [+ EF]
*SUB: en <dan hij> [//] dan Paddington+Bear pakt de hartjes <uit de> [//] uit het oven.
*SUB: en dan eet [/] eet hij de koekjes. [+ EF]

@Situation: Donald Duck and the bird
*SUB: Donald [/] Donald ziet een vogel. [+ EF]
*SUB: hij [/] hij gooit de vogel weg.
%com: target: "slaan" (or some such verb) instead of "weggooien"
*SUB: en dan de vogel <vlo:Gt@u> [//] <vluGt@u> [//] vlucht naar een stoel.
*SUB: en dan is de vogel heel # +/.
*WIL: ++ plat.
*SUB: +, plat. [+ EF]
*SHA: en waarom heeft ie (he)m weggeslaan dan?
*SUB: omdat de vogel zit op zijn koffie. [+ ell]

@Situation: Dikkie Dik and the chimney pot
*SUB: de poes klimt uit de raam. [+ EF]
*SUB: hij klimt op de [/] ## +/.
*SHA: ++ dak.
*SUB: +, <de dak> [/] de dak. [+ EF]

Example transcript of coded proficiency data (cont'd)

*SUB: <en nog een> [///] en hij loopt voor een lange tijd.
%com: target: without "voor"
*SUB: en dan komt hij naar een [/] +/.
*WIL: ++ een schoorsteen.
*SUB: +, een schoorsteen.
%com: target: "naar" should be "bij"
*SUB: <en hij klimt> [//] <en hij gaat> [/] en hij gaat op.
*WIL: en dan?
*SUB: gaat hij in.
*SUB: en dan ziet hij uit als een zwarte piets.

@Situation: Boxes
*SUB: een meisje doet een doos in een doos. [+ EF]
*SUB: dan [/] <en dan> [/] dan een meisje zit in een doos.
*SUB: en ze +... [+ inc]
*SHA: wat is de doos geworden?
*SUB: een boot. [+ EF] [+ ell]
*SUB: ze gaat in één boot. [+ EF]
*SUB: and +... [+ inc]
*SHA: wat zijn dit?
*SUB: tennis +... [+ inc] [+ ell]
*SHA: het is min of meer hetzelfde als in het engels.
*SUB: tennis+rackets. [+ EF] [+ ell]
*SUB: dan gaat de jongen een beer <in één> [///] op de bed staan en een doos naar +... [+ inc]
*SUB: dan knipt de jongen uit.
%com: target: with object
*SUB: knip [//] knipt [//] de jongen knipt de papier.
%com: target: "knipt in het papier"
*SUB: en dan <kijkt het als een ##> [/] +/.
*SUB: +, kijkt het als een kroon.
%com: target: "kijkt het als" should be "ziet het eruit als"
*WIL: wie heeft zo (ee)n hoed op?
*SUB: een koning. [+ EF] [+ ell]
*WIL: een koning.
*WIL: dan is (he)t een kroon.
*SUB: de kroon. [+ R]
%com: order of utterances changed so that CLAN counts utterance ending with +/. and utterance beginning with +, as one
*SUB: dan een vogel hij doet een doos op zijn hoofd. [+ EF]
*SUB: dan is ie [//] zijn hoofd een televisie. [+ EF]

@End

Table A. L1 children: Picture description task data

Subject	No. of utterances	Verbal density score (average number of finite and non-finite verbs per T-unit)	Lexical diversity score (V/√N)
7-year-olds			
7#1	41	1.07	6.62
7#2	58	1.21	6.71
7#3	53	1.36	6.13
7#4	44	1.48	5.97
7#5	37	1.38	5.77
7#6	40	1.25	5.25
7#7	31	1.42	5.01
7#8	65	1.19	5.62
7#9	66	1.18	7.44
7#10	39	1.36	5.88
9-year-olds			
9#1	55	1.55	6.48
9#2	38	1.24	6.49
9#3	54	1.32	6.17
9#4	55	1.46	6.72
9#5	57	1.65	6.17
9#6	43	1.30	6.19
9#7	36	1.36	4.97
9#8	40	1.45	6.04
9#9	77	1.43	7.30
9#10	58	1.36	6.14
11-year-olds			
11#1	54	1.35	5.97
11#2	51	1.37	5.55
11#3	36	1.31	6.05
11#4	39	1.23	5.79
11#5	49	1.51	5.93
11#6	42	1.45	6.14
11#7	37	1.46	5.61
11#8	44	1.46	6.06
11#9	47	1.45	5.25
11#10	**35**	1.40	5.65

Table B. L1 adults: Picture description task data

Subject	No. of utterances	Verbal density score (average number of finite and non-finite verbs per T-unit)	Lexical diversity score (V/√N)
Native adults			
N#1	49	1.61	7.33
N#2	52	1.46	6.43
N#3	50	1.56	6.52
N#4	60	1.88	7.88
N#5	76	1.38	8.21
N#6	47	1.32	6.07
N#7	62	1.97	7.20
N#8	82	1.46	6.61
N#9	53	1.40	6.58
N#10	61	1.75	7.87

Table C. L2 children: Proficiency data for production experiment (Chapter 5)

Subject	No. of utterances	Verbal density score		Lexical diversity score		Rate of errorfree utterances	Prof. score	Prof. level
		Average no. finite and non-finite verbs per T-unit	Z-score	V/√N	Z-score			
C1P	10	1.10	-1.11	5.17	-0.05	40%	-0.60	Low
C3P	12	1.08	-1.22	2.41	-1.99	8%	-2.06	Low
C4P	18	1.06	-1.39	3.84	-0.99	33%	-1.23	Low
C5P	25	1.20	-0.49	3.67	-1.11	20%	-1.19	Low
C6P	18	1.28	0.00	3.18	-1.45	28%	-1.01	Low
C8P	29	1.00	-1.74	3.84	-0.99	21%	-1.59	Low
C9P	31	1.19	-0.53	4.72	-0.37	26%	-0.78	Low
C10P	46	1.28	0.03	4.78	-0.33	11%	-0.83	Low
C11P	41	1.27	-0.07	5.37	0.08	68%	0.37	Mid
C12P	19	1.37	0.56	4.07	-0.83	21%	-0.66	Low
C13P	25	1.88	1.48	5.48	-0.65	44%	0.18	Mid
C14P	39	1.21	-0.46	5.37	0.08	39%	-0.32	Mid
C15P	51	1.35	0.46	7.38	1.49	69%	1.18	High
C16P	48	1.65	2.29	5.96	0.50	69%	1.43	High
C17P	81	1.16	-0.74	5.71	0.32	70%	0.26	Mid
C18P	38	1.13	-0.91	3.94	-0.92	50%	-0.71	Low
C19P	52	1.27	-0.06	5.87	0.43	21%	-0.34	Mid
C20P	33	1.30	0.15	5.80	0.38	61%	0.44	Mid
C21P	58	1.47	1.17	5.19	-0.04	57%	0.57	High
C22P	46	1.57	1.78	7.45	1.54	78%	1.87	High
C23P	45	1.49	1.31	6.71	1.02	42%	0.81	High
C24P	50	1.38	0.63	7.04	1.25	84%	1.42	High
C25P	84	1.45	1.08	8.47	2.25	89%	2.12	High

Table D. L2 adults: Proficiency data for production experiment (Chapter 5)

Subject	No. of utterances	Verbal density score		Lexical diversity score		Rate of errorfree utterances	Prof. score	Prof. level
		Average no. finite and non-finite verbs per T-unit	Z-score	V/√N	Z-score			
A1P	37	1.11	-1.06	4.06	-0.83	43%	-0.86	Low
A2P	30	1.37	0.55	5.71	0.32	30%	0.00	Mid
A3P	41	1.22	-0.37	5.92	0.46	46%	0.02	Mid
A4P	30	1.30	0.13	4.93	-0.22	23%	-0.51	Low
A5P	30	1.77	1.11	5.51	-0.64	40%	-0.02	Mid
A6P	43	1.63	0.66	7.28	0.39	61%	0.63	High
A7P	51	1.96	1.74	6.88	0.16	59%	0.90	High
A8P	54	1.74	1.03	8.04	0.83	57%	0.90	High
A9P	76	1.32	-0.36	7.76	0.67	76%	0.66	High
A10P	23	1.35	-0.25	5.11	-0.87	56%	-0.33	Mid
A11P	35	1.09	-1.11	5.15	-0.85	37%	-0.99	Low
A12P	28	1.00	-1.39	4.66	-1.13	61%	-0.78	Low
A13P	61	1.41	-0.05	8.50	1.10	72%	0.89	High
A14P	44	1.09	-1.09	5.23	-0.80	55%	-0.65	Low
A15P	65	1.49	0.22	8.41	1.05	39%	0.35	Mid
A16P	47	.94	-1.60	3.68	-1.70	34%	-1.60	Low
A17P	82	1.51	0.28	8.86	1.30	78%	1.21	High
A18P	61	1.49	0.22	9.57	1.72	64%	1.11	High
A19P	79	1.66	0.76	8.45	1.07	76%	1.25	High
A20P	36	1.19	-0.76	5.28	-0.77	17%	-1.20	Low
A21P	42	1.05	-1.23	4.98	-0.95	52%	-0.80	Low
A22P	66	1.53	0.34	6.72	0.06	65%	0.46	Mid

Table E. L2 children: Proficiency data for comprehension experiment (Chapter 6)

Subject	No. of utterances	Verbal density score		Lexical diversity score		Rate of errorfree utterances	Prof. score	Prof. level
		Average no. finite and non-finite verbs per T-unit	Z-score	V/√N	Z-score			
C1C	64	1.63	1.60	7.54	2.04	55%	1.42	High
C3C	51	1.16	-0.67	3.74	-1.48	24%	-1.52	Low
C4C	36	1.42	0.59	3.69	-1.53	14%	-1.25	Low
C6C	40	1.20	-0.46	3.99	-1.26	45%	-0.95	Low
C7C	47	1.38	0.43	6.26	0.85	62%	0.60	High
C8C	50	.74	-2.70	3.73	-1.50	40%	-1.98	Low
C9C	70	1.41	0.58	6.23	0.82	67%	0.74	High
C10C	44	1.43	0.66	5.63	0.26	77%	0.72	High
C11C	37	1.22	-0.39	4.72	-0.58	43%	-0.66	Low
C12C	51	1.39	0.47	7.24	1.75	69%	1.13	High
C13C	23	1.13	-0.80	3.98	-1.26	39%	-1.19	Low
C14C	49	1.41	0.55	6.07	0.68	71%	0.75	High
C15C	52	1.19	-0.50	5.23	-0.11	60%	-0.20	Mid
C16C	49	1.06	-1.14	3.58	-1.64	43%	-1.40	Low
C17C	41	1.17	-0.60	3.99	-1.26	42%	-1.07	Low
C18C	54	1.02	-1.34	5.17	-0.16	61%	-0.51	Low
C19C	45	1.58	1.37	4.18	-1.08	64%	0.17	Mid
C20C	75	1.32	0.12	6.26	0.85	68%	0.60	High
C21C	46	1.37	0.36	5.84	0.46	74%	0.63	High
C22C	47	1.43	0.64	6.41	0.99	60%	0.70	High
C23C	57	1.16	-0.67	4.64	-0.65	26%	-1.11	Low
C24C	44	1.41	0.55	4.92	-0.40	55%	-0.03	Mid
C25C	58	1.50	0.99	7.02	1.56	78%	1.41	High
C26C	53	1.43	0.67	7.11	1.64	87%	1.49	High
C27C	48	1.02	-1.33	5.03	-0.29	13%	-1.45	Low
C28C	50	1.12	-0.85	5.83	0.45	60%	-0.08	Mid
C29C	63	1.14	-0.74	5.68	0.32	57%	-0.15	Mid
C30C	49	1.51	1.04	5.65	0.28	86%	1.02	High
C31C	36	1.47	0.86	5.47	0.12	83%	0.84	High
C32C	46	1.41	0.57	5.98	0.59	72%	0.72	High
C33C	47	1.40	0.53	6.23	0.82	87%	1.09	High
C34C	57	1.49	0.95	6.96	1.50	97%	1.71	High
C35C	46	1.02	-1.33	4.75	-0.55	24%	-1.36	Low
C36C	63	1.37	0.34	4.89	-0.42	71%	0.19	Mid

Table F. L2 adults: Proficiency data for comprehension experiment (Chapter 6)

Subject	No. of utterances	Verbal density score		Lexical diversity score		Rate of errorfree utterances	Prof. score	Prof. level
		Average no. finite and non-finite verbs per T-unit	Z-score	V/√N	Z-score			
A1C	64	1.41	0.54	5.10	-0.22	17%	-0.64	Low
A2C	52	1.29	-0.04	6.26	0.85	64%	0.46	Mid
A3C	46	1.11	-0.91	4.26	-1.01	28%	-1.31	Low
A4C	45	1.22	-0.36	4.42	-0.86	20%	-1.20	Low
A5C	52	1.31	0.06	5.82	0.44	52%	0.11	Mid
A10C	41	.95	-1.67	3.70	-1.53	24%	-1.89	Low
A12C	36	1.56	1.27	5.49	0.13	44%	0.28	Mid
A13C	51	1.33	-0.18	7.64	0.49	86%	0.67	High
A14C	71	1.30	-0.31	7.34	0.29	55%	-0.04	Mid
A15C	77	1.83	1.61	7.15	0.17	48%	0.50	Mid
A16C	25	1.12	-0.94	3.96	-1.89	40%	-1.49	Low
A17C	91	1.36	-0.07	7.28	0.25	50%	-0.07	Mid
A18C	50	1.10	-1.01	4.10	-1.80	52%	-1.26	Low
A19C	44	1.41	0.10	5.90	-0.64	46%	-0.46	Mid
A20C	61	1.57	0.69	8.80	1.23	69%	0.99	High
A21C	42	1.36	-0.09	5.86	-0.66	45%	-0.55	Low
A22C	39	1.49	0.38	6.26	-0.41	69%	0.17	Mid
A23C	43	1.72	1.22	7.24	0.23	86%	1.07	High
A24C	125	1.59	0.75	8.26	0.89	85%	1.16	High
A25C	58	1.05	-1.19	5.24	-1.06	60%	-0.85	Low
A26C	72	1.51	0.47	7.77	0.57	81%	0.84	High
A27C	64	.91	-1.71	8.94	1.32	95%	0.62	High
A28C	117	1.88	1.79	8.61	1.11	70%	1.37	High
A29C	65	1.00	-1.37	5.27	-1.04	65%	-0.83	Low
A30C	76	1.34	-.14	8.36	0.95	79%	0.75	High
A32C	46	1.24	-0.27	4.82	-0.49	26%	-0.89	Low
A33C	51	1.02	-1.34	4.88	-0.43	61%	-0.63	Low
A34C	26	1.00	-1.44	4.08	-1.17	46%	-1.25	Low
A35C	61	1.66	1.75	6.74	1.30	75%	1.53	High
A36C	57	1.68	1.89	5.88	0.50	60%	0.95	High
A37C	79	1.33	0.16	6.07	0.67	86%	0.87	High

APPENDIX C

PRODUCTION EXPERIMENT (CHAPTER 5)

Experimental items: Definite condition

A1	*lezen*
Context:	Koekiemonster heeft vandaag niets te doen. Hij verveelt zich en kijkt of er iets leuks te doen is.
Koekiemonster:	Kijk, zegt Koekiemonster, een boek! Ik haat boeken. Dus dat ga ik NIET lezen.
Karel de Kikker:	Koekiemonster gaat het boek lezen.
Experimenter:	Nee hé? Vertel jij maar aan Karel de Kikker wat er echt gaat gebeuren.
Targetlike:	Koekiemonster gaat het boek NIET lezen.
Non-targetlike:	*Koekiemonster gaat NIET het boek lezen.

A1	*read*
Context:	Cookie Monster doesn't have anything to do today. He's bored and looking to see if there's anything nice to do.
Cookie Monster:	Look, says CookieMonster, a book! I hate books. So I'm NOT going to read it.
Karel de Kikker:	Cookie Monster will read the book.
Experimenter:	No, hey? You tell Charles the Frog what'll really happen.
Subject:	Cookie Monster will NOT read the book.

A2	*plukken*
Context:	Nijntje wil bloemen plukken voor haar moeder. Ze heeft al een heleboel bloemen. Ze vindt nog een bloem.
Nijntje:	Kijk, zegt Nijntje, een bloem. Die vind ik te groot. Dus die ga ik NIET plukken.
Karel de Kikker:	Nijntje gaat de bloem plukken.
Experimenter:	Nee, hé? Wat gebeurt er echt?
Targetlike:	Nijntje gaat de bloem NIET plukken.
Non-targetlike:	* Nijntje gaat NIET de bloem plukken.

A2	*pick*
Context:	Miffy wants to pick flowers for her mother. She already has a lot of flowers. She finds another flower.
Miffy:	Look, says Miffy, a flower. I think it's too big. So I'm NOT going to pick it.
Karel de Kikker:	Miffy will pick the flower.
Experimenter:	No, hey? You tell Charles the Frog what'll really happen.
Subject:	Miffy will NOT pick the flower.

A3	*vangen*
Context:	Goofy heeft een nieuwe hengel gekregen. Hij wil er iets mee vangen. Maar er zwemmen geen vissen in de vijver. Er ligt een laars in de vijver.
Goofy:	Kijk, zegt Goofy, een laars. Laarzen kun je niet eten. Dus die ga ik NIET vangen.
Karel de Kikker:	Goofy gaat de laars vangen.
Experimenter:	Nee, hé? Wat gebeurt er echt?
Targetlike:	Goofy gaat de laars NIET vangen.
Non-targetlike:	* Goofy gaat NIET de laars vangen.

A3	*catch*
Context:	Goofy has got a new fishing rod. He wants to catch something with it. But there aren't any fish swimming in the pond. There's a boot in the pond.
Goofy:	Look, says Goofy, a boot. You can't catch boots. So I'm NOT going to catch it.
Karel de Kikker:	Goofy will catch the boot.
Experimenter:	No, hey? You tell Charles the Frog what'll really happen.
Subject:	Goofy will NOT catch the boot.

A4	*opeten*
Context:	Paddington heeft honger. Hij wil iets eten.
Beertje Padding-ton:	Kijk, zegt ie, een boterham met kaas. Bah, dat lust ik niet. Dus dat ga ik NIET opeten.
Karel de Kikker:	Paddington gaat de boterham opeten.
Experimenter:	Nee hé? Wat gebeurt er echt?
Targetlike:	Paddington gaat de boterham NIET opeten.
Non-targetlike:	* Paddington gaat NIET de boterham opeten.

A4	eat (up)
Context:	Paddington is hungry. He wants to eat something.
Paddington Bear:	Look, he says, a cheese sandwich. Urgh, I don't like them. So I'm NOT going to eat it.
Karel de Kikker:	Paddington will eat the sandwich..
Experimenter:	No, hey? You tell Charles the Frog what'll really happen.
Subject:	Paddington will NOT eat the sandwich.

A5	*uitknippen*
Context:	Bert is aan het knutselen. Hij wil iets uitknippen.
Bert:	Kijk, zegt ie, een teddybeer. Nou, die vind ik helemaal niet mooi. Dus die ga ik NIET uitknippen.
Karel de Kikker:	Bert gaat de teddy beer uitknippen.
Experimenter:	Nee hé? Wat gebeurt er echt?
Targetlike:	Bert gaat de teddybeer NIET uitknippen.
Non-targetlike:	* Bert gaat NIET de teddybeer uitknippen.

A5	*cut out*
Context:	Bert is making things. He wants to cut something out.
Bert:	Look, he says, a teddy bear. Now, I don't think that's very nice at all. So I'm not going to cut it out.
Karel de Kikker:	Bert will NOT cut out the teddy bear.
Experimenter:	No, hey? You tell Charles the Frog what'll really happen.
Subject:	Bert will NOT cut out the teddy bear.

A6	*natekenen*
Context:	Ernie wil iets natekenen.
Ernie:	Kijk, een giraffe. Wat een grote giraffe. Hij is echt te groot. Dus die ga ik NIET natekenen.
Karel de Kikker:	Ernie gaat de giraffe natekenen.
Experimenter:	Nee hé? Wat gebeurt er echt?
Targetlike:	Ernie gaat de giraffe NIET natekenen.
Non-targetlike:	* Ernie gaat NIET de giraffe natekenen.

A6	*copy*
Context:	Ernie wants to copy something.
Ernie:	Look, he says, a giraffe. What a big giraffe. It's really too big. So I'm NOT going to copy it.
Karel de Kikker:	Ernie will copy the giraffe.
Experimenter:	No, hey? You tell Charles the Frog what'll really happen.
Subject:	Ernie will NOT copy the giraffe.

Experimental items: Specific indefinite condition

C1	*lezen*
Ernie:	Kijk, boeken. Grote boeken, kleine boeken. Ik heb zin om te lezen. Maar ik heb niet zo veel tijd. Dus een groot boek ga ik NIET lezen.
Karel de Kikker:	Oh, ik hoorde je niet goed. WAT gaat Ernie niet doen?
Targetlike:	Ernie gaat een (groot) boek niet lezen.
Non-targetlike:	* Ernie gaat niet een / geen boek lezen.

C1	*read*
Ernie:	Look, books. Big books, little books. I fancy reading something. But I don't have very much time. So I won't read a big book.
Karel de Kikker:	Oh, I didn't hear you very well. WHAT won't Ernie do?
Subject:	Ernie won't read a (big) book

C2	*plukken*
Mickey Mouse:	Mickey staat in de tuin. Kijk, zegt ie, wat een mooie bloemen, zeg. Gele bloemen, rode bloemen. Ik houd van bloemen. Ik ga ze plukken. Maar dan is de tuin leeg. Dus een rode bloem ga ik NIET plukken.
Karel de Kikker:	Ik heb niet goed verstaan. WAT gaat Mickey Mouse niet doen?
Targetlike:	Mickey Mouse gaat een (rode) bloem niet plukken.
Non-targetlike:	* Mickey Mouse gaat niet een / geen (rode) bloem plukken.

C2	*pick*
Mickey Mouse:	Mickey's in the garden. Look, he says, what beautiful flowers. Yellow flowers, red flowers. I love flowers. I'm going to pick them. But then the garden will be empty. So I won't pick a red flower.
Karel de Kikker:	I didn't understand. WHAT won't Mickey do?
Subject:	Mickey Mouse won't pick a (red) flower

C3	*vangen*
Bert:	Kijk, zegt Bert, vissen. Rode vissen, oranje vissen. Ik ga ze vangen. Maar dan is de vijver leeg. Dus een oranje vis ga ik niet vangen.
Karel de Kikker:	Ik heb niet goed verstaan. WAT gaat Bert niet doen?
Targetlike:	Bert gaat een (oranje) vis NIET vangen.
Non-targetlike:	* Bert gaat niet een / geen vis vangen.

C3	*catch*
Bert:	Look, says Bert, fish. Red fish, orange fish. I'm going to catch them. But then the pond will be empty. So I won't catch an orange fish.
Karel de Kikker:	I didn't understand. WHAT won't Bert do?
Subject:	Bert won't catch an (orange) fish.

C4	*opeten*
Winnie de Poeh:	Kijk, zegt Winnie de Poeh, cakejes. Gele cakejes, oranje cakejes. Ik ga ze opeten. Maar als ik ze allemaal opeet, word ik misselijk. Dus een oranje cakeje ga ik NIET opeten.
Karel de Kikker:	Ik snap het niet. WAT gaat Winnie de Poeh niet doen?
Targetlike:	Winnie de Poeh gaat een (oranje) cakeje niet opeten.
Non-targetlike:	* Winnie de Poeh gaat niet een / geen cakeje opeten.

C4	*eat (up)*
Winnie the Pooh:	Look, says Winnie the Pooh, cakes. Yellow cakes, orange cakes. I'm going to eat them. But if I eat them all, I'll be sick. So I won't eat an (orange) cake.
Karel de Kikker:	I don't get it. WHAT won't Winnie the Pooh do?
Subject:	Winnie the Pooh won't eat an (orange) cake.
C5	*uitknippen*
Beertje Padding-ton:	Paddington heeft zin om iets uit te knippen. Kijk, zegt Paddington, mooie hartjes. Er zijn roze hartjes en blauwe hartjes. Uitknippen vind ik héél leuk. Maar het is zonde ze allemaal in een keer uit te knippen. Dus een blauw hartje ga ik NIET uitknippen.
Karel de Kikker:	Ik hoorde je niet goed. WAT gaat Paddington niet doen?
Targetlike:	Paddington gaat een (blauw) plaatje niet uitknippen.
Non-targetlike:	* Paddington gaat niet een / geen plaatje uitknippen.
C5	*cut out*
Paddington Bear:	Paddington fancies cutting something out. Look, says Paddington,beautiful hearts. There are pink hearts and blue hearts. Cutting out is very nice. But it's a shame to cut them all out in one go. So I won't cut out a (blue) heart.
Karel de Kikker:	I didn't hear you very well. WHAT won't Paddington do?
Subject:	Paddington won't cut out a (blue) heart
C6	*natekenen*
Donald Duck:	Donald Duck is aan het tekenen. Kijk, zegt ie, een boek met mooie plaatjes. Plaatjes van olifanten en plaatjes van slangen. Ik wil al die plaatjes natekenen. Maar ik heb niet genoeg papier. Dus een plaatje van een olifant ga ik NIET natekenen.
Karel de Kikker:	Ik zat niet op te letten. WAT gaat Donald Duck niet doen?
Targetlike:	Donald Duck gaat een plaatje (van een olifant) niet natekenen.
Non-targetlike:	* Donald Duck gaat niet een / geen plaatje natekenen.
C6	*copy*
Donald Duck:	Donald Duck is drawing. Look, he says, a book with beautiful pictures. Pictures of elephants and pictures of snakes. I want to copy all those pictures. But I don't have enough paper. So I won't copy a picture of an elephant.
Karel de Kikker:	I wasn't paying attention. WHAT won't Donald Duck do?
Subject:	Donald Duck won't cut out (a picture of) an elephant.

Experimental items: Non-specific indefinite condition

B1	*lezen*
Beertje Paddington:	Paddington loopt op straat. Kijk, zegt ie, de bibliotheek. Daar vind je heel veel boeken. Ik zou daar heel graag binnen gaan en een boek gaan lezen. Maar ik ben mijn bril vergeten. Dus dat ga ik NIET doen.
Karel de Kikker:	Oh, ik heb het niet goed verstaan. WAT gaat Paddington niet doen?
Targetlike:	Paddington gaat niet een boek / geen boek(en) lezen.
Non-targetlike:	Paddington gaat een boek niet lezen.
B1	*read*
Paddington Bear:	Paddington is walking about. Look, he says, the library. You find lots of books there. I'd really like to go inside and read a book. But I've forgotten my glasses. So I won't do that.
Karel de Kikker:	Oh, I didn't understand. WHAT won't Paddington do?
Subject:	Paddington won't read a book.
B2	*plukken*
Goofy:	Kijk, zegt Goofy, wat een mooie tuin. Er zijn veel bloemen in die tuin. Ik houd van bloemen. Ik zou graag een bloem plukken maar dat mag niet zo maar. Dus dat ga ik NIET doen.
Karel de Kikker:	Oh, ik hoorde je niet goed. WAT gaat Goofy niet doen?
Targetlike:	Goofy gaat niet een bloem / geen bloem(en) plukken.
Non-targetlike:	Goofy gaat een bloem niet plukken.
B2	*pick*
Goofy:	Look, says Goofy, what a beautiful garden. T here are lots of flowers in that garden. I love flowers. I'd like to pick a flower but you're not allowed to do that. So I won't do that.
Karel de Kikker:	Oh, didn't hear you very well. WHAT won't Goofy do?
Subject:	Goofy won't pick a flower.
B3	*vangen*
Ernie:	Ik heb zin om iets te vangen, zegt Ernie. Kijk, een vijver. Er zitten kikkers in. Maar een kikker is moeilijk om te vangen. Dus dat ga ik NIET doen.
Karel de Kikker:	Oh, ik hoorde je niet goed. WAT gaat Ernie niet doen?
Targetlike:	Ernie gaat niet een kikker / geen kikker(s) vangen.
Non-targetlike:	Ernie gaat een kikker niet vangen.
B3	*catch*
Ernie:	I fancy catching something, says Ernie. Look, a pond. There are frogs in it. But a frog is difficult to catch. So I won't do that.
Karel de Kikker:	Oh, I didn't hear you very well. WHAT won't Ernie do?
Subject:	Ernie won't catch a frog.

B4	*opeten*
Knorretje:	Mmmm, zegt Knorretje, ik heb zin om iets te eten. Ik kan een snoepje opeten maar dat is slecht voor je tanden. Dus dat ga ik NIET doen.
Karel de Kikker:	Oh, ik zat niet op te letten. WAT gaat Knorretje niet doen?
Targetlike:	Knorretje gaat niet een snoepje / geen snoepje eten.
Non-targetlike:	Knorretje gaat een snoepje niet eten.

B4	*eat (up)*
Piglet:	Mmmm, says Piglet, I fancy eating something. I could eat a sweetie but that's bad for your teeth. So I won't do that.
Karel de Kikker:	Oh, I wasn't paying attention. WHAT won't Piglet do?
Subject:	Piglet won't eat a sweetie.

B5	*uitknippen*
Koekie-monster:	Mmm, zegt Koekiemonster, ik heb zin om iets uit te knippen. Kijk, bomen, heel veel bomen. Ik kan een boom uitknippen maar dat vind ik te moeilijk. Dus dat ga ik NIET doen.
Karel de Kikker:	Oh, ik heb het niet goed verstaan. WAT gaat Koekiemonster niet doen?
Targetlike:	Koekiemonster gaat niet een boom / geen boom (uit)knippen.
Non-targetlike:	Koekiemonster gaat een boom niet (uit)knippen.

B5	*cut out*
Cookie Monster:	Mmm, says Cookie Monster, I fancy cutting out something. Look, trees, lots of trees. I could cut out a tree but I find that too difficult. So I won't do that.
Karel de Kikker:	Oh, I didn't understand very well. WHAT won't Cookie Monster do?
Subject:	Cookie Monster won't cut out a tree.

B6	*natekenen*
Nijntje:	Mmmm, zegt Nijntje, ik heb zin om iets na te tekenen. Dat zijn veel vogels. Wat een mooie vogels. Ik wil een vogel natekenen maar dat is wel heel moeilijk. Ik denk niet dat ik dat kan. Dus dat ga ik NIET doen.
Karel de Kikker:	Oh, ik snap het niet. WAT gaat Nijntje niet doen?
Targetlike:	Nijntje gaat niet een vogel / geen vogel natekenen.
Non-targetlike:	Nijntje gaat een vogel niet natekenen.

B6	*copy*
Miffy:	Mmmm, says Miffy, I fancy copying something. That's a lot of flowers. What beautiful flowers. I want to copy a bird but that's really difficult. I don't think I can. So I won't do that.
Karel de Kikker:	Oh, I don't understand. WHAT won't Miffy do?
Subject:	Miffy won't copy a bird.

Fillers with similar format to indefinite conditions

D1

Winnie de Poeh:	Mickey Mouse is jarig vandaag. Dus, zegt Winnie de Poeh, ik ga voor hem een cadeautje kopen.
Karel de Kikker:	Oh ik snap het niet. WAT gaat Winnie de Poeh doen?
Target:	Winnie de Poeh gaat een cadeautje voor Mickey Mouse kopen.

D1

Winnie the Pooh:	It's Mickey Mouse's birthday. So, Winnie the Pooh, I'm going to buy him a present.
Karel de Kikker:	Oh I don't get it. WHAT will Winnie the Pooh do?
Target:	Winnie the Pooh will buy a present for Mickey Mouse.

D2

Nijntje:	Nijntje wil met Bert gaan spelen. Dus zij gaat hem opbellen.
Karel de Kikker:	Oh ik zat niet op te letten. WAT gaat Nijntje doen?
Target:	Nijntje gaat Bert opbellen.

D2

Miffy:	Miffy will go and play with Bert. So she's going to call him.
Karel de Kikker:	Oh I wasn't paying attention. WHAT will Miffy do?
Target:	Miffy will call Bert.

Fillers taken from Schaeffer's original experiment

E1

Bert:	Kijk, zegt Bert, een cakeje! Mmmm, dat ga ik eens lekker opeten. En dat ga ik HIER doen.
Karel de Kikker:	Het cakeje gaat Bert HIER opeten.
Target:	Ja, goed zo!

E1

Bert:	Look, says Bert, a cake! Mmmm, I'm going to eat that. And I'll do that HERE.
Karel de Kikker:	Bert will eat the cake HERE.
Target:	Yes, well done!

E2

Nijntje:	Hé, zegt Nijntje, wat een mooie olifant. Die ga ik natekenen. Maar dat ga ik STRAKS pas doen.
Karel de Kikker:	De olifant gaat Nijntje STRAKS natekenen.
Target:	Ja, goed zo!

E2

Nijntje:	Hey, says Miffy, what a beautiful elephant! I'm going to copy that. But I'll do that LATER.
Karel de Kikker:	Miffy will copy the elephant LATER.
Target:	Yes, well done!

Fillers on binding

Target: Yes-response (where the protagonist is carrying out an action to him- or herself)

G1

Context:	Hier hebben we twee katten– hier en hier – en een muis, een lamp en een tafel.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	De kat naast de muis krabt zich.
Target:	Ja, goed zo.

G1

Context:	Here we have two cats – here and here – and a mouse, a lamp and a table.
Experimenter:	Karel, can YOU tell us something about this picture?
Karel de Kikker:	The cat next to the mouse is scratching itself.
Target:	Yes, well done.

G2

Context:	Hier hebben we twee soldaten – hier en hier – en een voetballer, in het park.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	De soldaat naast de voetballer krabt zich.
Target:	Ja, goed zo.

G2

Context:	Here we have two soldiers – here en here – and a footballer, in the park.
Experimenter:	Karel, can YOU tell us something about this picture?
Karel de Kikker:	The soldier next to the footballer himself.
Target:	Yes, well done.

G3

Context:	Hier hebben we twee baby aapjes – hier en hier – en een moederaap, in de dierentuin.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	Het babyaapje naast de moederaap wast zich.
Target:	Ja, goed zo.

G3

Context:	Here we have two baby monkeys – here and here – and a mother monkey, in the zoo.
Experimenter:	Karel, can YOU tell us something about this picture?
Karel de Kikker:	The baby monkey next to the mother monkey is washing itself.
Target:	Yes, well done.

G4

Context:	Hier hebben we twee jongens – hier en hier – en een man, op de kermis.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	Het jongetje nast de man verstopt een lolly achter zich.
Target:	Ja, goed zo.

G4

Context:	Here we have two boys – here and here – and a man, at the fair.
Experimenter:	Karel, can YOU tell us something about this picture?
Karel de Kikker:	The boy next to the man is hiding a lolly behind himself.
Target:	Yes, well done.

G5

Context:	Hier hebben we twee meisjes – hier en hier – en een vrouw, en het regent.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	Het meisje naast de vrouw houdt een paraplu boven zich.
Target:	Ja, goed zo.

G5

Context:	Here we have two girls – here and here – and a woman, and it's raining.
Experimenter:	Karel, can YOU tell us something about this picture?
Karel de Kikker:	The girl next to the woman is holding an umbrella above herself.
Target:	Yes, well done.

G6

Context:	Hier hebben we twee dokters – hier en hier – en een politieagent, en twee stoelen.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	De dokter naast de politieagent legt een kussen onder zich.
Target:	Ja, goed zo.

G6

Context:	Here we have two doctors – here and here – and a policeman, and two chairs.
Experimenter:	Karel, can YOU tell us something about this picture?
Karel de Kikker:	The doctor next to the policeman is putting a cushion under himself.
Target:	Yes, well done.

Target: No-response (where the protagonist is carrying out an action to the object embedded in the PP subject)

G7

Context:	Hier hebben we twee beertjes – hier en hier – en een giraffe, in het bos.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	De beer naast de giraffe krabt zich.
Target:	Nee, de beer krabt de giraffe.

G7

Context:	Here we have two bears – here and here – and a giraffe, in the wood.
Experimenter:	Karel, can YOU tell us something about this picture?
Karel de Kikker:	The bear next to the giraffe is scratching itself.
Target:	No, the bear's scratching the giraffe.

G8

Context:	Hier hebben we twee boerinnen – hier en hier, en een prinses, in de badkamer.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	De boerin naast de prinses wast zich.
Target:	Nee, de boerin wast de prinses.
	*Ja, goed zo.

G8

Context:	Here we have two lady-farmers – here and here – and a princess, in the bathroom.
Experimenter:	Karel, can you tell us something about this picture?
Karel de Kikker:	The lady-farmer next to the princess is washing himself.
Target:	No, the lady-farmer is washing the princess.

G9

Context:	Hier hebben we twee koningen – hier en hier – en een boer, bij het kasteel.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	De koning naast de boer krabt zich.
Target:	Nee, de koning krabt de boer.
	*Ja, goed zo.

G9

Context:	Here we have two kings – here en here – and a farmer, by the castle.
Experimenter:	Karel, can you tell us something about this picture?
Karel de Kikker:	The king next to the farmer is scratching himself.
Target:	No, the king is scratching the farmer

G10

Context:	Hier hebben we twee meisjes – hier en hier – en een vrouw, op de kermis.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	Het meisje naast de vrouw vertstopt een pop achter zich.
Target:	Nee, het meisje verstopt een pop achter de vrouw.
	*Ja, goed zo.

G10

Context:	Here we have two girls – here and here – and a woman, at the fair.
Experimenter:	Karel, can you tell us something about this picture?
Karel de Kikker:	The girl next to the woman is hiding a doll behind herself.
Target:	No, the girl is hiding a doll behind the woman.

G11

Context:	Hier hebben we twee olifanten – hier en hier – en een beer in het bos, en het regent.
Experimenter:	Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker:	De olifant naast de beer houdt een paraplu boven zich.
Target:	Nee, de olifant houdt een paraplu boven de beer.

G11
Context: Here we have two elephants – here and here – and a bear in the wood, and it's raining.
Experimenter: Karel, can you tell us something about this picture?
Karel de Kikker: The elephant next to the bear is holding an umbrella above himself.
Target: No, the elephant is holding an umbrella above the bear.

G12
Context: Hier hebben we twee meisjes – hier en hier – en een oma, en een stoel.
Experimenter: Karel, kun JIJ ons iets vertellen over dit plaatje?
Karel de Kikker: Het meisje naast de oma legt een kussen onder zich.
Target: Nee, het meisje legt een kussen onder de oma.

G12
Context: Here we have two girls– here and here – and a granny, and a chair.
Experimenter: Karel, can you tell us something about this picture?
Karel de Kikker: The girl next to the granny is putting a cushion under herself.
Target: No, the girl is putting a cushion under the granny.

Remaining fillers

F1
Koekiemonster: Koekiemonster is jarig. Hij geeft een feestje. Hij is heel blij.
Karel de Kikker: Oh, ik begrijp het. Koekiemonster is jarig.
Target: Ja, goed zo!

F1
Cookie Monster: It's Cookie Monster's birthday. He's giving a party. He's really happy.
Karel de Kikker: Oh, I understand. It's Cookie Monster's birthday.
Target: Yes, well done!

F2
Knorretje: Knorretje is ziek. Hij is verkouden. Hij is niet blij.
Karel de Kikker: Oh, ik zie het al. Knorretje is ziek.
Target: Ja, goed zo.

F2
Piglet: Piglet is ill. He's got a cold. He's not happy.
Karel de Kikker: Oh, I see it already. Piglet is ill.
Target: Yes, well done!

F4
Winnie de Poeh: Winnie de Poeh heeft niet genoeg geslapen. Hij is dood moe.
Karel de Kikker: Oh, ik begrijp het. Winnie de Poeh heeft héél veel energie. Daarom gaat ie niet slapen.
Experimenter: Nee, he? Vertel jij aan Karel de Kikker wat er aan de hand is!
Target: Winnie de Poeh is moe.

F4

Winnie the Pooh:	Winnie the Pooh hasn't slept enough. He's exhausted.
Karel de Kikker:	Oh, I understand. Winnie the Pooh has lots of energy. And so he's not going to sleep.
Experimenter:	No, hey? You tell Charles the Frog what'll really happen.
Target:	Winnie the Pooh is tired.

F5

Bert:	Bert heeft zijn pak aan. De zon schijnt heel hard. Dus hij heeft het veel te warm.
Karel de Kikker:	Oh, ik zie het al. Bert heeft het koud. Daarom heeft ie zijn pak aan.
Experimenter:	Nee, he? Vertel jij aan Karel de Kikker wat er aan de hand is!
Target:	Bert heeft het warm.

F5

Bert:	Bert is wearing his suit. The sun's shining really hard. So he's far too warm.
Karel de Kikker:	Oh, I see it already. Bert's cold.That's why he's got his suit on.
Experimenter:	No, hey? You tell Charles the Frog what'll really happen.
Target:	Bert's warm.

Table A. L1 adults (production). Individual results: Percentage and number of objects scrambled

Subject	Definite condition (Target: Scrambled)				Specific indefinite condition (Target: Scrambled)					Non-specific indefinite condition (Target: Non-scrambled)				
	%	n	Other	Excl.	%	n	Other	Excl.	*geen*-forms	%	n	Other	Excl.	*geen*-forms
N#1P	100	5/5	1		100	3/3	3			0	0/6			6/6
N#2P	100	6/6			80	4/5		1	1/5	0	0/6			6/6
N#3P	100	6/6			75	3/4	1	1		0	0/4	2		3/4
N#4P	100	3/3	3		100	6/6				0	0/1	4	1	1/1
N#5P	83.3	5/6			100	5/5		1		0	0/2	3	1	1/2
N#6P	100	5/5		1	100	5/5		1		0	0/6			6/6
N#7P	100	6/6			100	5/5		1		0	0/5	1		5/5
N#8P	100	6/6			100	2/2	4			0	0/3	3		3/3
N#9P	100	6/6			100	6/6				0	0/4	2		4/4
N#10P	100	6/6			100	1/1	4	1		0	0/3	3		3/3
N#11P	100	5/5	1		66.7	2/3	2	1		0	0/1	4	1	1/1

Table B. L2 children (production): Biodata

Subject	Country of origin	Age at time of testing	Age at first exp. (years; months)	Length of exp. (years; months)	Contact with Dutch[a]	Other languages
C1P	USA	5;3.25	4	1	Mod.	
C2P	USA	6;2.14	6	0.2	Ltd.	
C3P	USA	6;4.3	5;2	1;2	Ltd.	
C4P	UK	7;5.18	4	3	Mod.	
C5P	USA	7;6.20	5	2	Mod.	
C6P	USA	7;9.7	5	2	Ltd.	
C7P	USA	7;9.13	5	2	Ltd.	
C8P	UK	8;3.24	5	3	Mod.	
C9P	UK	9;4.0	6	3	Mod.	
C10P	USA	9;7.20	7	2	Mod.	
C11P	USA	10;0.6	7	3	Mod.	
C12P	Australia	10;3,17	7	3	Ltd.	
C13P	UK	17;4,5	4	13	Mod.	? School French
C14P	UK	10;2.16	6;9	3;5	Mod.	
C15P	Canada	11;2,8	6;2	6;0	Mod.	
C16P	UK	11;0,3	6;0	5;0	Ext.	
C17P	USA	8;9,18	4;0	4;9	Mod.	
C18P	UK	8;5,11	5;10	2;7	Mod.	
C19P	UK	9;10,9	7;1[b]	2;9	Mod.	
C20P	Israel	11;3,8	5	6;3	Ltd.	School French for 1 yr
C21P	UK	8;8,21	4	4;4	Ext.	
C22P	UK	10;10,30	6;6	4;4	Mod.	
C23P	Canada	9;3,13	4;4	4;11	Mod.	
C24P	Canada	8;8,26	5;8[c]	3;0	Ext.	
C25P	USA	10;4,5	5;5	4;11	Ext.	School French for 1 yr

[a] Limited (Ltd.) contact with Dutch means that the subject's only contact with Dutch is at school (children) or foreign language classes (adults) and possibly in basic interactions in shops, etc. Moderate (Mod.) contact means that the subject has some Dutch-speaking friends and/or contact with Dutch neighbours. Extensive (Ext.)contact means that subjects meet one or more of the following criteria: for the adults – the subject's partner is a native-speaker of Dutch and interaction with him or her is predominanly in Dutch and/or the subject works in a Dutch-speaking environment (at least some of the time); for the children – the subject has attended a Dutch-speaking school; for either children or adults – the subject lives with one or more native-speakers who speak Dutch to him or her.

[b] This subject was born in The Netherlands but has only ever attended an English-speaking school and nursery. His first exposure was therefore in Dutch language classes at school.

[c] This subject was in The Netherlands for two months prior to first exposure.

Table C. L2 adults (production): Biodata

Subject	Country of origin	Age at time of testing	Age at first exp. (years; months)	Length of exp. (years; months)	Contact with Dutch[a]	Other languages
A1P	USA	10;0,12	8	2	Mod.	
A2P	UK[b]	10;1,18	8	2	Mod.	
A3P	Nigeria	12;4,27	8	4	Mod.	
A4P	UK	13;5,6	12	1;6	Mod.	School German
A5P	UK	15;8,24	8	7	Ltd.	? School French
A6P	Australia	18;5.8	12	6	Ext.	? School French
A7P	UK	23;6.21	23	0;6	Ext.	School French
A8P	UK	24;7.15	24	1;6	Ext.	School French
A9P	UK	25;2.25	21	3;6	Ext.	
A10P	UK	25;6.11	24	1	Ltd.	
A11P	UK	27;9.4	25	2;2	Mod.	School French and Spanish
A12P	Canada	28;6.5	27	1	Ltd.	Fluent French
A13P	UK	28;7.28	22	6;6	Ext.	School French and German
A14P	USA	29;1.25	28	1;3	Mod.	Spanish, 1year Finnish
A15P	New Zealand	32;6.6	28	3;10	Ext.	School French and German
A16P	UK	33;0.16	32	1;3	Ltd.	
A17P	UK	37;9.21	26	11	Ext.	German A-level
A18P	UK	38;6.25	24	14	Ext.	
A19P	UK	50;6.30	23	27	Ext.	Intermediate French and German; can read Russian
A20P	UK	10;10,0	8;5	2;5	Mod.	
A21P	USA	10;3,0	8;10	1;5	Mod.	
A22P	USA[c]	10;6,8	10;3	0;3	Ext.	
A23P	UK	32;2,2	32	0;5	Ltd.	Degree in French and Arabic; beginners Spanish

[a] See Table B for details.
[b] This subject has a Dutch-speaking mother, but she spoke English to him (while living in the UK). His first real contact with Dutch was therefore upon arrival in The Netherlands
[c] This subject has a Dutch-speaking mother, but she spoke English to him (while living in the USA). His first real contact with Dutch was therefore upon arrival in The Netherlands.

Table D. L2 children (production): Individual results in ascending order of proficiency. Percentage and number of targetlike responses.

Subj.	Prof. score	Definite condition (Target: Scrambled)						Specific indefinite condition (Target: Scrambled)					
		%	n	SVO	Other	Excl.	Resp.	%	n	SVO	Other	Excl.	Resp.
L2 CHILDREN: LOW PROFICIENCY													
C2P	---	0	0/4	3	1		-	---			4		
C7P	---	0	0/5	5		1	-	0	0/1	1	4	1	-
C3P	-2.058	0	0/2	2		1	-	0	0/1			3	-
C8P	-1.586	0	0/3			1	-	---			4		
C4P	-1.226	0	0/4				-	0	0/3		1		-
C5P	-1.189	0	0/6	1			-	0	0/3		1	1	-
C6P	-1.013	0	0/4	4			-	0	0/3	3	1		-
C10P	-0.826	40	2/5	3	1		±	0	0/4	4	2		-
C9P	-0.776	16.7	1/6				±	33.3	1/3		3[a]		±
C18P	-0.713	100	6/6				+	100	5/5		1		+
C12P	-0.659	75	3/4				±	66.7	2/3		3		±
C1P	-0.596	33.3	1/3				±	33.3	1/3				±
L2 CHILDREN: MID PROFICIENCY													
C19P	-0.339	100	5/5		1		+	80	4/5			1	±
C14P	-0.323	100	6/6				+	50	3/6				±
C13P	0.180	0	0/6				-	0	0/4		2		-
C17P	0.263	100	6/6				+	100	3/3		1	2	+
C11P	0.370	100	6/6				+	100	5/5				+
C20P	0.440	80	4/5		1		±	100	5/5		1		+
L2 CHILDREN: HIGH PROFICIENCY													
C21P	0.566	16.7	1/6				±	66.7	2/3		3[b]		±
C23P	0.811	100	6/6				+	100	4/4		2		+
C15P	1.185	100	2/2		4		+	100	1/1		5		+
C24P	1.425	100	3/3		3		+	100	6/6				+
C16P	1.435	100	5/5		1		+	100	6/6				+
C22P	1.873	100	5/5		1		+	100	6/6				+
C25P	2.124	100	6/6				+	75	3/4		2		±

GUIDE TO COLUMN HEADINGS:

SVO Number of utterances included in % which have Negation-Verb-Object order

Other Number of alternative utterances which did not include the constituents necessary to determine whether scrambling had taken place.

Excl Number of utterances which were excluded (e.g. because the subject's response was inappropriate or because s/he misunderstood the experimental scenario).

Resp Reponse pattern: "–" no scrambling, "±" optional scrambling, "+" obligatory scrambling

NB: Due to space constraints, Table D is split into two parts. The first presents the two conditions where scrambling is required (definite and specific indefinite). The second repeats the specific indefinite with the non-specific indefinite condition. Footnotes appear below.

Table D (cont'd). L2 children (production): Individual results in ascending order of proficiency. Ppercentage and number of targetlike responses.

Subj.	Prof. score	Specific indefinite condition (Target: Scrambled)						Non-specific indefinite condition (Target: Non-scrambled)					
		%	n	SVO	Other	Excl.	Resp.	%	n	SVO	Other	Excl.	Resp.
L2 CHILDREN: LOW PROFICIENCY													
C2P	---	---			4			---			3	1	
C7P	---	0	0/1	1	4	1	-	0	0/1	1	5		---
C3P	-2.058	0	0/1			3	-	100	2/2		2		-
C8P	-1.586	---			4			---			3	1	
C4P	-1.226	0	0/3		1		-	100	1/1		3		-
C5P	-1.189	0	0/3		1	1	-			1	4		---
C6P	-1.013	0	0/3	3	1		-	---			4		
C10P	-0.826	0	0/4	4	2		-	0	0/2	2	4		---
C9P	-0.776	33.3	1/3		3[a]		±	83.3	5/6		0		±
C18P	-0.713	100	5/5		1		+	0	0/2		4		+
C12P	-0.659	66.7	2/3		3		±	---			5		
C1P	-0.596	33.3	1/3				±	100	3/3		0		-
L2 CHILDREN: MID PROFICIENCY													
C19P	-0.339	80	4/5			1	±	0	0/4		2		+
C14P	-0.323	50	3/6				±	100	4/4		2		-
C13P	0.180	0	0/4		2		-	100	1/1		4	1	-
C17P	0.263	100	3/3		1	2	+	50	2/4		2		±
C11P	0.370	100	5/5				+	---			4	1	
C20P	0.440	100	5/5		1		+	100	2/2		4		-
L2 CHILDREN: HIGH PROFICIENCY													
C21P	0.566	66.7	2/3		3		±	100	2/2		4		-
C23P	0.811	100	4/4		2		+	100	1/1		5		-
C15P	1.185	100	1/1		5		+	---			6		
C24P	1.425	100	6/6				+	100	4/4		2		-
C16P	1.435	100	6/6				+	0	0/5		1[c]		+
C22P	1.873	100	6/6				+	100	2/2		3	1	-
C25P	2.124	75	3/4		2		±	100	5/5		1		-

[a] This includes one double negative, as in (i):

(i) Donald gaat niet een olifant niet natekenen
Donald goes not an elephant not copy
'Donald isn't going not to copy an elephant.'

[b] This includes two bare numerals (*één* 'one') in scrambled position.
[c] This utterance is a double negative.

Table E. L2 adults (production): Individual results in ascending order of proficiency. Percentage and number of targetlike responses.

Subj.	Prof. score	Definite condition (Target: Scrambled)						Specific indefinite condition (Target: Scrambled)					
		%	n	SVO	Other	Excl.	Resp.	%	n	SVO	Other	Excl.	Resp.
L2 ADULTS: LOW PROFICIENCY													
A23P	---	0	0/5	5	1		-	0	0/5	5		1	-
A16P	-1.599	0	0/6				-	0	0/1		5		-
A20P	-1.202	33.3	2/6				±	0	0/5				-
A11P	-0.991	0	0/6				-	---			6		
A1P	-0.857	0	0/3	1	1	1	-	0	0/3	1	3		-
A21P	-0.799	100	6/6				+	100	5/5				+
A12P	-0.784	0	0/6				-	---			6		
A14P	-0.646	20	1/5		1		±	0	0/1		4	1	-
A4P	-0.513	20	1/5	1			±	0	0/3	2	2	1	-
L2 ADULTS: MID PROFICIENCY													
A10P	-0.327	33.3	2/6				±	25	1/4		2[a]		±
A2P	0.002	100	4/4		2		+	25	1/4		2[a]		±
A3P	0.023	100	6/6				+	0	0/2		3	1	-
A5P	0.023	100	6/6				+	100	4/4		2		+
A15P	0.348	100	3/3		3		+	100	3/3		3		+
A22P	0.457	100	4/4		2		+	100	5/5		1		+
L2 ADULTS: HIGH PROFICIENCY													
A6P	0.631	100	4/4		1		+	100	4/4		2		+
A9P	0.665	100	6/6				+	100	1/1		5		+
A13P	0.888	100	2/2		4		+	100	3/3		2	1	+
A7P	0.902	0	0/6				-	0	0/2		4		-
A8P	0.902	100	5/5		1		+	---			5	1	
A18P	1.107	100	6/6				+	100	1/1		4	1	+
A17P	1.209	100	2/2		4		+	100	2/2		4		+
A19P	1.245	100	6/6				+	100	2/2		3	1	+

GUIDE TO COLUMN HEADINGS:

SVO Number of utterances included in % which have Negation-Verb-Object order
Other Number of alternative utterances which did not include the constituents necessary to determine whether scrambling had taken place.
Excl Number of utterances which were excluded (e.g. because the subject's response was inappropriate or because s/he misunderstood the experimental scenario).
Resp Reponse pattern: "–" no scrambling, "±" optional scrambling, "+" obligatory scrambling

NB: Due to space constraints, Table E is split into two parts. The first presents the two conditions where scrambling is required (definite and specific indefinite). The second repeats the specific indefinite with the non-specific indefinite condition. Footnotes appear below.

Table E (cont'd). L2 adults (production): Individual results in ascending order of proficiency. Percentage of targetlike responses.

Subj.	Prof. score	Specific indefinite condition (Target: Scrambled)						Non-specific indefinite condition (Target: Non-scrambled)					
		%	n	SVO	Other	Excl.	Resp.	%	n	SVO	Other	Excl.	Resp.
L2 ADULTS: LOW PROFICIENCY													
A23P	---	0	0/5	5		1	-	0	0/5	5		1	---
A16P	-1.599	0	0/1		5		-	---			6		
A20P	-1.202	0	0/5				-	100	1/1		4		-
A11P	-0.991	---			6			100	1/1		5		-
A1P	-0.857	0	0/3	1	3		-	0	0/1	1	3		---
A21P	-0.799	100	5/5				+	0	0/5		1		+
A12P	-0.784	---			6			100	2/2		4		-
A14P	-0.646	0	0/1		4	1	-	---			6		
A4P	-0.513	0	0/3	2	2	1	-	---			5	1	
L2 ADULTS: MID PROFICIENCY													
A10P	-0.327	25	1/4		2[a]		±	100	3/3		3		-
A2P	0.002	25	1/4		2[a]		±	100	1/1		5		-
A3P	0.023	0	0/2		3	1	-	100	5/5			1	-
A5P	0.023	100	4/4		2		+	---			6		
A15P	0.348	100	3/3		3		+	100	3/3		3		-
A22P	0.457	100	5/5		1		+	33.3	1/3		3		±
L2 ADULTS: HIGH PROFICIENCY													
A6P	0.631	100	4/4		2		+	---			6		
A9P	0.665	100	1/1		5		+	100	2/2		3	1	-
A13P	0.888	100	3/3		2	1	+	100	5/5		1		-
A7P	0.902	0	0/2		4		-	100	4/4		2		-
A8P	0.902	---			5	1		100	1/1		5		-
A18P	1.107	100	1/1		4	1	+	100	6/6				-
A17P	1.209	100	2/2		4		+	100	1/1		5		-
A19P	1.245	100	2/2		3	1	+	100	3/3		3		-

[a] Both these utterances are double negatives (see above for example).

Table F. L2 children (production): Use of één inspecific indefinite condition (in ascending order of proficiency)

Subj.	Prof. score	Specific indefinite condition (percentage of scrambled objects)				Number of indefinite determiners realised as *één*		
		%	n	SVO	Resp.	Neg-Verb-Obj	Neg-Obj-Verb	Obj-Neg-Verb
L2 CHILDREN: LOW PROFICIENCY								
C2P	---	---						
C7P	---	0	0/1	1	-	0/1	---	---
C3P	-2.058	0	0/1		-	---	0/1	---
C8P	-1.586	---				---	---	---
C4P	-1.226	0	0/3		-	---	2/3	---
C5P	-1.189	0	0/3		-	---	2/3	---
C6P	-1.013	0	0/3	3	-	3/3	---	---
C10P	-0.826	0	0/4	4	-	3/4	---	---
C9P	-0.776	33.3	1/3		±	---	0/2	0/1
C18P	-0.713	100	5/5		+	---	---	1/5
C12P	-0.659	66.7	2/3		±	---	0/1	1/2
C1P	-0.596	33.3	1/3		±	---	2/2	1/1
L2 CHILDREN: INTERMEDIATE PROFICIENCY								
C19P	-0.339	80	4/5		±	---	1/1	0/4
C14P	-0.323	50	3/6		±	---	2/3	1/3
C13P	0.180	0	0/4		-	---	1/4	---
C17P	0.263	100	3/3		+	---	---	3/3
C11P	0.370	100	5/5		+	---	---	0/5
C20P	0.440	100	5/5		+	---	---	4/5
L2 CHILDREN: HIGH PROFICIENCY								
C21P	0.566	66.7	2/3		±	---	1/1	2/2
C23P	0.811	100	4/4		+	---	---	4/4
C15P	1.185	100	1/1		+	---	---	1/1
C24P	1.425	100	6/6		+	---	---	6/6
C16P	1.435	100	6/6		+	---	---	3/6
C22P	1.873	100	6/6		+	---	---	1/6
C25P	2.124	75	3/4		±	---	0/1	2/3

Table F. L2 adults (production): Use of één inspecific indefinite condition (in ascending order of proficiency)

Subj.	Prof. score	Specific indefinite condition (percentage of scrambled objects)				Number of indefinite determiners realised as *één*		
		%	n	SVO	Resp.	Neg-Verb-Obj	Neg-Obj-Verb	Obj-Neg-Verb
L2 ADULTS: LOW PROFICIENCY								
A23P	---	0	0/5	5	-	3/5	---	---
A16P	-1.599	0	0/1		-	---	0/1	---
A20P	-1.202	0	0/5		-	---	5/5	---
A11P	-0.991	---						
A1P	-0.857	0	0/3	1	-	0/1	1/2	---
A21P	-0.799	100	5/5		+	---	---	0/5
A12P	-0.784	---						
A14P	-0.646	0	0/1		-	---	0/1	---
A4P	-0.513	0	0/3	2	-	0/2	0/1	---
L2 ADULTS: INTERMEDIATE PROFICIENCY								
A10P	-0.327	25	1/4		±	---	0/3	0/1
A2P	0.002	25	1/4		±	---	0/3	1/1
A3P	0.023	0	0/2		-	---	1/2	---
A5P	0.023	100	4/4		+	---	---	4/4
A15P	0.348	100	3/3		+	---	---	3/3
A22P	0.457	100	5/5		+	---	---	3/5
L2 ADULTS: HIGH PROFICIENCY								
A6P	0.631	100	4/4		+	---	---	3/4
A9P	0.665	100	1/1		+	---	---	1/1
A13P	0.888	100	3/3		+	---	---	3/3
A7P	0.902	0	0/2		-	---	0/2	---
A8P	0.902	---						
A18P	1.107	100	1/1		+	---	---	0/1
A17P	1.209	100	2/2		+	---	---	2/2
A19P	1.245	100	2/2		+	---	---	2/2

APPENDIX D

COMPREHENSION EXPERIMENTS (CHAPTER 6)

Truth value judgement task : Warm-up items

w1 *schommel*
Hier is een meisje. Ze wil denk ik gaan schommelen.
Hier schommelt ze.
En hier schommelt ze.
En nu gaat ze weer weg.
Het meisje heeft geschommeld. (accept)

w1 *swing*
Here's a girl. I think she wants to go on the swing.
Here she's swinging.
And here she's swinging.
And now she's leaving.
The girl has swung. (accept)

w2 *tekening*
Hier is een meisje. Ze wil denk ik gaan tekenen.
Hier tekent ze.
En hier tekent ze nog meer.
En nu gaat ze weer weg.
Het meisje heeft niet getekend. (reject)

w2 *drawing*
Here's a girl. I think she wants to draw.
Here she's drawing.
And here she's still drawing.
And now she's leaving.
The girl has not drawn. (reject)

w3 *slaap*
Hier is een jongen. Hij wil denk ik gaan slapen.
Hier slaapt hij.
Hier staat hij weer op.
En nu gaat hij weer weg.
De jongen heeft niet geslapen. (reject)

w3 *sleep*
Here's a boy. I think he wants to go to sleep.
Here he's sleeping.
Here he's getting up again.
And now he's leaving.
The boy did not sleep. (reject)

Truth value judgement task (*twee keer*): Experimental items

t1	*een bal gooien*
	Dit is een meisje. En dit zijn ballen. Ze wil ze denk ik gooien.
	Hier gooit ze een bal.
	En hier gooit ze een bal.
	En nu gaat ze weer weg.
Scrambling	Het meisje heeft een bal twee keer gegooid. (reject)
No scrambling	Het meisje heeft twee keer een bal gegooid. (accept)
t1	*throw a ball*
	This is a girl. These are balls. I think she wants to throw them.
	Here she's throwing a ball.
	And here she's throwing a ball.
	And now she's leaving.
Scrambling	The girl threw a (certain) ball twice. (reject)
No scrambling	The girl threw a(ny) ball twice. (accept)
t2	*een teddybeer zoenen*
	Dit is een meisje. En dit zijn teddy-beren. Ze wil ze denk ik zoenen.
	Hier zoent ze een teddy-beer.
	En hier zoent ze een teddy-beer.
	En nu gaat ze weer weg.
Scrambling	Het meisje heeft een teddy-beer twee keer gezoend. (reject)
No scrambling	Het meisje heeft twee keer een teddy-beer gezoend. (accept)
t2	*kiss a teddy bear*
	This is a girl. And these are teddy bears. I think she wants to kiss them.
	Here she's kissing a teddy bear.
	And here she's kissing a teddy bear.
	And now she's leaving.
Scrambling	The girl kissed a (certain) teddy bear twice. (reject)
No scrambling	The girl kissed a(ny) teddy bear twice. (accept)
t3	*een aap kietelen*
	Dit is een meisje. En dit zijn apen. Zij wil ze denk ik kietelen.
	Hier kietelt ze een aap.
	En hier kietelt ze een aap.
	En nu gaat ze weer weg.
Scrambling	Het meisje heeft een aap twee keer gekieteld. (reject)
No scrambling	Het meisje heeft twee keer een aap gekieteld. (accept)
t3	*tickle a monkey*
	This is a girl. And these are monkeys. I think she wants to tickle them.
	Here she's tickling a monkey.
	And here she's tickling a monkey.
	And now she's leaving.
Scrambling	The girl tickled a (certain) monkey twice. (reject)
No scrambling	The girl tickled a(ny) monkey twice. (accept)

t4	*een varken knijpen*
	Dit is een jongen. En dit zijn varkens. Hij wil ze denk ik knijpen.
	Hier knijpt ie een varken.
	En hier knijpt ie een varken.
	En nu gaat ie weer weg.
Scrambling	De jongen heeft een varken twee keer geknepen. (reject)
No scrambling	De jongen heeft twee keer een varken geknepen. (accept)

t4	*pinch a pig*
	This is a boy. And these are pigs. I think he wants to pinch them.
	Here he's pinching a pig.
	And here he's pinching a pig.
	And now he's leaving.
Scrambling	The boy pinched a (certain) pig twice. (reject)
No scrambling	The boy pinched a(ny) pig twice. (accept)

t5	*een hond aaien*
	Dit is een jongen. En dit zijn honden. Hij wil ze denk ik aaien.
	Hier aait ie een hond.
	En hier aait ie een hond.
	En nu gaat ie weer weg.
Scrambling	De jongen heeft een hond twee keer geaaid. (reject)
No scrambling	De jongen heeft twee keer een hond geaaid. (accept)

t5	*stroke a dog*
	This is a boy. And these are dogs. I think he wants to stroke them.
	Here he's stroking a dog.
	And here he's stroking a dog.
	And now he's leaving.
Scrambling	The boy stroked a (certain) dog twice. (reject)
No scrambling	The boy stroked a(ny) dog twice. (accept)

Truth value judgement task (*twee keer*): Fillers

Target: No-response

f1	*een frisbee gooien*
	Dit is een jongen. Dit is een bal en dit zijn frisbees. Hij wil ze denk ik gooien.
	Hier gooit ie de bal.
	En hier gooit ie een frisbee.
	En nu gaat ie weer weg.
Scrambling	De jongen heeft een frisbee twee keer gegooid. (reject)
No scrambling	De jongen heeft twee keer een frisbee gegooid. (reject)

f1 *throw a frisbee*
This is a boy. This is a ball and these are frisbees. I think he wants to throw them.
Here he's throwing the ball.
And here he's throwing a frisbee.
And now he's leaving.
Scrambling The boy threw a (certain) frisbee twice. (reject)
No scrambling The boy threw a(ny) frisbee twice. (reject)

f2 *een teddybeer zoenen*
Dit is een jongen. Dit is een teddy-beer en dit zijn meisjes. Hij wil ze denk ik zoenen.
Hier zoent ie de teddy-beer.
En hier zoent ie een meisje.
En nu gaat ie weer weg.
Scrambling De jongen heeft een meisje twee keer gezoend. (reject)
No scrambling De jongen heeft twee keer een meisje gezoend. (reject)

f2 *kiss a teddy bear*
This is a boy. This is a teddy bear. And these are girls. I think he wants to kiss them.
Here he's kissing the teddy bear.
And here he's kissing a girl.
And now he's leaving.
Scrambling The boy kissed a (certain) girl twice. (reject)
No scrambling The boy kissed a(ny) girl twice. (reject)

f3 *een hond kietelen*
Dit is een jongen. Dit is een aap en dit zijn honden. Hij wil ze denk ik kietelen.
Hier kietelt ie de aap.
En hier kietelt ie een hond.
En nu gaat ie weer weg.
Scrambling De jongen heeft een hond twee keer gekieteld. (reject)
No scrambling De jongen heeft twee keer een hond gekieteld. (reject)

f3 *tickle a dog*
This is a boy. This is a monkey and these are dogs. I think he wants to tickle them.
Here he's tickling the monkey.
And here he's tickling a dog.
And now he's leaving.
Scrambling The boy tickled a (certain) dog twice. (reject)
No scrambling The boy tickled a(ny) dog twice . (reject)

f4	*een poes knijpen*
	Dit is een meisje. Dit is een varken en dit zijn poezen. Zij wil ze denk ik knijpen.
	Hier knijpt ze het varken.
	En hier knijpt ze een poes.
	En nu gaat ze weer weg.
Scrambling	Het meisje heeft een poes twee keer geknepen. (reject)
No scrambling	Het meisje heeft twee keer een poes geknepen. (reject)

f4	*pinch a cat*
	This is a girl. This is a pig and these are cats. I think she wants to pinch them.
	Here she's pinching the pig.
	And here she's pinching a cat.
	And now she's leaving.
Scrambling	The girl pinched a (certain) cat twice. (reject)
No scrambling	The girl pinched a(ny) cat twice. (reject)

f5	*een paard aaien*
	Dit is een meisje. Dit is een hond en dit zijn paarden. Zij wil ze denk ik aaien.
	Hier aait ze de hond.
	En hier aait ze een paard.
	En nu gaat ze weer weg.
Scrambling	Het meisje heeft een paard twee keer geaaid. (reject)
No scrambling	Het meisje heeft twee keer een paard geaaid. (reject)

f5	*stroke a horse*
	This is a girl. This is a dog and these are horses. I think she wants to stroke them.
	Here she's stroking the dog.
	And here she's stroking a horse.
	And now she's leaving.
Scrambling	The girl stroked a (certain) horse twice. (reject)
No scrambling	The girl stroked a(ny) horse twice. (reject)

Target: Yes-response

f6	*het pijltje gooien*
	Dit is een meisje. En dit is een pijltje. Zij wil het denk ik gooien.
	Hier gooit ze het pijltje.
	En hier gooit ze het pijltje.
	En nu gaat ze weer weg.
Scrambling	Het meisje heeft het pijlje twee keer gegooid. (accept)
No scrambling	Het meisje heeft twee keer het pijltje gegooid. (accept)

f6	*throw the dart*
	This is a girl. And this is a dart. I think she wants to throw it.
	Here she's throwing the dart.
	And here she's throwing the dart.
	And now she's leaving.
Scrambling	The girl threw the dart twice. (accept)
No scrambling	The girl threw the dart twice. (accept)
f7	*het konijn zoenen*
	Dit is een meisje. En dit is een konijn. Zij wil het denk ik zoenen.
	Hier zoent ze het konijn.
	En hier zoent ze het konijn.
	En nu gaat ze weer weg.
Scrambling	Het meisje heeft het konijn twee keer gezoend. (accept)
No scrambling	Het meisje heeft twee keer het konijn gezoend. (accept)
f7	*kiss the rabbit*
	This is a girl. And this is a rabbit. I think she wants to kiss it.
	Here she's kissing thea rabbit.
	And here she's kissing the rabbit.
	And now she's leaving.
Scrambling	The girl kissed the rabbit twice. (accept)
No scrambling	The girl kissed the rabbit twice. (accept)
f8	*het muisje kietelen*
	Dit is een jongen. En dit is een muisje. Hij wil het denk ik kietelen.
	Hier kietelt ie het muisje.
	En hier kietelt ie het muisje.
	En nu gaat ie weer weg.
Scrambling	De jongen heeft het muisje twee keer gekieteld. (accept)
No scrambling	De jongen heeft twee keer het muisje gekieteld. (accept)
f8	*tickle the mouse*
	This is a boy. And this is a mouse. I think he wants to tickle it.
	Here he's tickling the mouse.
	And here he's tickling the mouse.
	And now he's leaving
Scrambling	The boy tickled the mouse twice. (accept)
No scrambling	The boy tickled the mouse twice. (accept)
f9	*het schaap knijpen*
	Dit is een jongen. En dit is een schaap. Hij wil het denk ik knijpen.
	Hier ie knijpt ie het schaap.
	En hier ie het schaap.
	En nu gaat ie weer weg.
Scrambling	De jongen heeft het schaap twee keer geknepen. (accept)
No scrambling	De jongen heeft twee keer het schaap geknepen. (accept)

f9	*pinch the sheep*
	This is a boy. And this is a sheep. I think he wants to pinch it.
	Here he's pinching the sheep.
	And here he's pinching the sheep.
	And now he's leaving.
Scrambling	The boy pinched the sheep twice. (accept)
No scrambling	The boy pinched the sheep twice. (accept)

f10	*het konijn aaien*
	Dit is jongen. En dit is een konijn. Hij wil het denk ik aaien.
	Hier aait ie het konijn.
	En hier aait ie het konijn.
	En nu gaat ie weer weg.
Scrambling	De jongen heeft het konijn twee keer geaaid. (accept)
No scrambling	De jongen heeft twee keer het konijn geaaid. (accept)

f10	*stroke the rabbit*
	This is a boy. And this is a rabbit. I think he wants to stroke them.
	Here he's stroking the rabbit.
	And here he's stroking the rabbit.
	And now he's leaving.
Scrambling	The boy stroked the rabbit twice. (accept)
No scrambling	The boy stroked the rabbit twice. (accept)

Truth value judgement task (negation): Experimental items

t1	*een vis vangen*
	Dit is een jongen. En dit zijn vissen. Die wil ie denk ik gaan vangen.
	Hier vangt ie een vis.
	En hier vangt ie een vis.
	En nu gaat ie weer weg.
	De jongen heeft een vis niet gevangen. (accept)

t1	*catch a fish*
	This is a boy. And these are fish. I think he wants to catch them.
	Here he's catching a fish.
	And here he's catching a fish.
	And now he's leaving.
	The boy hasn't caught a (certain) fish. (accept)

t2	*een appel plukken*
	Dit is een meisje en dit zijn appels. Die wil ze denk ik plukken.
	Hier plukt ze een appel.
	En hier plukt ze een appel.
	En nu gaat ze weer weg.
	Het meisje heeft een appel niet geplukt. (accept)

t2 *pick an apple*
This is a girl and these are apples. I think she wants to pick them.
Here she's picking an apple.
And here she's picking an apple.
And now she's leaving.
The girl hasn't picked a (certain) apple. (accept)

t3 *een ketting stelen*
Dit is een boef. En dit zijn kettingen. Die wil ie denk ik stelen.
Hier steelt ie een ketting.
En hier steelt ie een ketting.
En nu gaat ie weer weg.
De boef heeft een ketting niet gestolen. (accept)

t3 *stealing a necklace*
This is a thief. And these are necklaces. I think he wants to steal them.
Here he's stealing a necklace.
And here he's stealing a necklace.
And now he's leaving.
The thief did not steal a (certain) necklace. (accept)

t4 *een bloes strijken*
Dit is een vrouw. En dit zijn bloezen. Die wil ze denk ik strijken.
Hier strijkt ze een bloes.
En hier strijkt ze een bloes.
En nu gaat ze weer weg.
De vrouw heeft een bloes niet gestreken. (accept)

t4 *ironing a blouse*
This is a woman. And these are blouses. I think she wants to iron them. ik strijken.
Here she's ironing a blouse.
And here she's ironing a blouse.
And now she's leaving.
The woman did not iron a (certain) necklace. (accept)

t5 *een koekje pakken*
Dit is een meisje. En dit zijn koekjes. Die wil ze denk ik pakken.
Hier pakt ze een koekje.
En hier pakt ze een koekje.
En nu gaat ze weer weg.
Het meisje heeft een koekje niet gepakt. (accept)

t5 *take a biscuit*
This is a girl. And these are biscuits. I think she wants to take them.
Here she's taking a biscuit.
And here she's taking a biscuit.
And now she's leaving.
The girl did not take a (certain) biscuit. (accept)

Truth value judgement task (negation): Fillers

Target: No-response

fh2 *hoed* (story depicts two hats)
Dit is een jongen. En dit zijn hoeden. Die wil ie denk ik opzetten.
Hier zet ie een hoed op.
En hier zet ie een hoed op.
En nu gaat ie weer weg.
De jongen heeft een hoed niet opgezet. (reject)

fh2 *hat* (story depicts two hats)
This is a boy. And these are hats. I think he wants to put them on.
Here he is putting a hat on.
And here he is putting a hat on.
And now he's leaving.
The boy didn't put a (certain) hat on. (reject)

fh4 *boek* (story depicts one book)
Dit is een meisje. En dit is een boek. Dat wil ze denk ik gaan lezen.
Hier leest ze…
… en hier leest ze.
En nu gaat ze weer weg.
Het meisje heeft een boek niet gelezen. (reject)

fh4 *book* (story depicts one book)
This is a girl. And this is a book. I think she want to read it.
Here she is reading…
… and here she's reading.
And now she's leaving.
The girl didn't read a (certain) book. (reject)

fh6 *kaars* (story depicts two candles)
Dit is een jongen. Dit zijn kaarsen. Ze zijn omgevallen, kijk maar. De jongen wil ze denk ik rechtop gaan zetten.
Hier zet ie een kaars rechtop.
En hier zet ie een kaars rechtop.
En nu gaat ie weer weg.
De jongen heeft een kaars niet rechtop gezet. (reject)

fh6 *candle* (story depicts two candles)
This is a boy. And these are candles. Look, they dropped. I think the boy wants to put them upright again.
Here he is putting a candle upright.
And here he's putting a candle upright.
And now he's leaving.
The boy didn't put a (certain) candle upright. (reject)

fh7 *sok* (story depicts one sock)
Dit is een man. En dit is een sok. Die wil ie denk ik wassen.
Hier wast ie de sok.
En hier wast ie de sok.
En nu gaat ie weer weg.
De man heeft een sok niet gewassen. (reject)

fh7 *sock* (story depicts one sock)
This is a man. And this is a sock. I think he wants to wash it.
Here he's washing the sock.
And here he's washing the sock.
And now he's leaving.
The man didn't wash a (certain) sock. (reject)

fh8 *appel* (story depicts one apple)
Dit is een meisje. En dit is een appel. Die wil ze denk ik opeten.
Hier eet ze.
En hier eet ze.
En nu gaat ze weer weg.
Het meisje heeft een appel niet gegeten. (reject)

fh8 *apple* (story depicts one apple)
This is a girl. And this is an apple. I think she wants to eat it.
Here she's eating.
And here she's eating.
And now she's leaving.
The girl didn't eat a (certain) apple. (reject)

Target: Yes-response

fh1 *wasgoed* (story depicts one T-shirt)
Dit is een jongen. En dit is een T-shirt. Dat wil ie denk ik ophangen.
Maar, hier houdt ie een T-shirt vast…
…en hier laat ie het op de grond vallen.
En nu gaat ie weer weg.
De jongen heeft een T-shirt niet opgehangen. (accept)

fh1 *laundry* (story depicts one T-shirt)
This is a boy. And this is a T-shirt. I think he wants to hang it up.
But, here he's holding a T-shirt...
…and here he drops it on the floor.
And now he's leaving.
The boy did not hang up a (certain) T-shirt. (accept)

fh3 *afwas* (story depicts one plate)
Dit is een vrouw. En dit is een bord. Dat wil ze denk ik afwassen.
Ze pakt het bord…
…en zet het op de aanrecht
En nu gaat ze weer weg.
De vrouw heeft een bord niet afgewassen. (accept)

fh3 *washing up* (story depicts one plate)
This is a woman. And this is a plate. I think she wants to clean it.
She is taking the plate…
…and she's placing it on the work-top.
And now she's leaving.
The woman didn't clean a (certain) plate. (accept)

fh5 *paard* (story depicts one horse)
Dit is een man met een borstel. En dit is een paard. Hij wil denk ik het paard gaan borstelen.
Hier kijkt de man naar het paard.
En hier kijkt ie naar het paard.
En nu gaat ie weer weg.
De man heeft een paard niet geborsteld. (accept)

fh5 *horse* (story depicts one horse)
This is a man with a brush. And this is a horse. I think he want to brush the horse.
Here the man is looking at the horse.
And here he is looking at the horse.
And now he's leaving.
The man didn't brush a (certain) horse. (accept)

Table A. L1 adults (comprehension): Individual results for truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Age at time of testing	Negation (target: yes)		*Twee keer*			
				Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
		%	n	%	n	%	n
CN#1	23	100%	(5/5)	100%	(5/5)	100%	(5/5)
CN#2	26	100%	(5/5)	100%	(5/5)	100%	(5/5)
CN#3[a]	21	100%	(5/5)	---	---	---	---
CN#4	18	100%	(5/5)	60%	(3/5)	100%	(5/5)
CN#5	43	100%	(5/5)	100%	(5/5)	60%	(3/5)
CN#6	24	---	---	100%	(5/5)	100%	(5/5)
CN#7	22	100%	(5/5)	100%	(5/5)	100%	(5/5)
CN#8	52	100%	(5/5)	0%	(0/5)	100%	(5/5)
CN#9	19	100%	(5/5)	100%	(5/5)	100%	(5/5)
CN#10	20	100%	(5/5)	100%	(5/5)	80%	(4/5)
CN#11	20	100%	(5/5)	100%	(5/5)	100%	(5/5)
CN#12	21	100%	(5/5)	100%	(5/5)	100%	(5/5)
CN#23	32	60%	(3/5)	100%	(5/5)	100%	(5/5)
CN#24	21	60%	(3/5)	100%	(5/5)	100%	(5/5)
CN#25	26	80%	(4/5)	100%	(5/5)	100%	(5/5)

[a] Subjects CN#3 and CN#6 were excluded from the twee keer and the negation tasks respectively because they misunderstood the task.

GUIDE TO CELL SHADING:

Black	Targetlike response pattern (acceptance on negation task and rejection on *twee keer* task
Grey	Mixed response pattern
White	Non-targetlike response pattern

Table B. L1 children (comprehension): Individual results on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Negation (target: yes)		*Twee keer*			
			Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
	%	n	%	n	%	n
L1 children: 7-year-olds						
7#1	60%	(3/5)	0%	(0/5)	100%	(5/5)
7#2	0%	(0/5)	0%	(0/5)	100%	(5/5)
7#3	0%	(0/5)	0%	(0/5)	100%	(5/5)
7#4	80%	(4/5)	100%	(5/5)	0%	(0/5)
7#5	0%	(0/5)	0%	(0/5)	100%	(5/5)
7#6	0%	(0/5)	100%	(5/5)	40%	(2/5)
7#7	0%	(0/5)	40%	(2/5)	100%	(5/5)
7#8	0%	(0/5)	0%	(0/5)	100%	(5/5)
7#9	0%	(0/5)	0%	(0/5)	100%	(5/5)
7#10	0%	(0/5)	0%	(0/5)	100%	(5/5)
7#11	0%	(0/5)	0%	(0/5)	100%	(5/5)
7#12	0%	(0/5)	0%	(0/5)	100%	(5/5)
7#13	0%	(0/5)	20%	(1/5)	100%	(5/5)
7#14	0%	(0/5)	100%	(5/5)	0%	(0/5)
L1 children: 8-year-olds						
8#1	0%	(0/5)	100%	(5/5)	100%	(5/5)
8#2	0%	(0/5)	0%	(0/5)	100%	(5/5)
8#3	0%	(0/5)	0%	(0/5)	100%	(5/5)
8#4	0%	(0/5)	0%	(0/5)	100%	(5/5)
8#5	0%	(0/5)	0%	(0/5)	100%	(5/5)
8#6	100%	(5/5)	0%	(0/5)	100%	(5/5)
8#7	100%	(5/5)	100%	(5/5)	40%	(2/5)
8#8	60%	(3/5)	0%	(0/5)	100%	(5/5)
8#9	0%	(0/5)	0%	(0/5)	100%	(5/5)
8#10	0%	(0/5)	0%	(0/5)	100%	(5/5)
8#11	0%	(0/5)	0%	(0/5)	100%	(5/5)
8#12	40%	(2/5)	0%	(0/5)	100%	(5/5)
8#13	0%	(0/5)	0%	(0/5)	100%	(5/5)
8#14	0%	(0/5)	80%	(4/5)	100%	(5/5)
8#15	no true negs		0%	(0/5)	100%	(5/5)
8#16	0%	(0/5)	0%	(0/5)	100%	(5/5)

Table B (cont'd). L1 children (comprehension): Individual results on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Negation (target: yes)		*Twee keer*			
			Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
	%	n	%	n	%	n
L1 children: 9-year-olds						
9#1	100%	(5/5)	60%	(3/5)	100%	(5/5)
9#2	100%	(5/5)	80%	(4/5)	80%	(4/5)
9#3	80%	(4/5)	100%	(5/5)	0%	(0/5)
9#4	0%	(0/5)	20%	(1/5)	100%	(5/5)
9#5	100%	(5/5)	40%	(2/5)	100%	(5/5)
9#6	100%	(5/5)	100%	(5/5)	80%	(4/5)
9#7	80%	(4/5)	100%	(5/5)	40%	(2/5)
9#8	40%	(2/5)	0%	(0/5)	100%	(5/5)
9#9	20%	(1/5)	0%	(0/5)	100%	(5/5)
9#10	0%	(0/5)	60%	(3/5)	100%	(5/5)
9#11	100%	(5/5)	100%	(5/5)	100%	(5/5)
9#12	60%	(3/5)	100%	(5/5)	100%	(5/5)
9#13	0%	(0/5)	0%	(0/5)	100%	(5/5)
9#14	100%	(5/5)	60%	(3/5)	100%	(5/5)
9#15	100%	(5/5)	20%	(1/5)	100%	(5/5)
L1 children: 11-year-olds						
11#1	0%	(0/5)	80%	(4/5)	100%	(5/5)
11#2	100%	(5/5)	100%	(5/5)	60%	(3/5)
11#3	100%	(5/5)	100%	(5/5)	100%	(5/5)
11#4	80%	(4/5)	100%	(5/5)	100%	(5/5)
11#5	100%	(5/5)	0%	(0/5)	100%	(5/5)
11#6	100%	(5/5)	0%	(0/5)	100%	(5/5)
11#7	80%	(4/5)	60%	(3/5)	80%	(4/5)
11#8	20%	(1/5)	100%	(5/5)	80%	(4/5)
11#9	80%	(4/5)	100%	(5/5)	0%	(0/5)
11#10	80%	(4/5)	60%	(3/5)	100%	(5/5)
11#11	100%	(5/5)	100%	(5/5)	100%	(5/5)
11#12	20%	(1/5)	100%	(5/5)	100%	(5/5)
11#13	0%	(0/5)	0%	(0/5)	100%	(5/5)
11#14	0%	(0/5)	0%	(0/5)	100%	(5/5)
11#15	0%	(0/5)	0%	(0/5)	100%	(5/5)

Table B (cont'd). L1 children (comprehension): Individual results on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Negation (target: yes)		*Twee keer*			
			Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
	%	n	%	n	%	n
L1 children: 12-year-olds						
12#1	100%	(5/5)	100%	(5/5)	100%	(5/5)
12#2	80%	(4/5)	100%	(5/5)	100%	(5/5)
12#3	80%	(4/5)	100%	(5/5)	100%	(5/5)
12#4	20%	(1/5)	100%	(5/5)	100%	(5/5)
12#5	100%	(5/5)	---	---	---	---
12#6	40%	(2/5)	20%	(1/5)	100%	(5/5)
12#7	no true negs		100%	(5/5)	100%	(5/5)
12#8	100%	(5/5)	100%	(5/5)	100%	(5/5)
12#9	100%	(5/5)	100%	(5/5)	100%	(5/5)
12#10	100%	(5/5)	100%	(5/5)	100%	(5/5)
12#11	100%	(5/5)	100%	(5/5)	40%	(2/5)
12#12	100%	(5/5)	100%	(5/5)	100%	(5/5)
12#13	80%	(4/5)	100%	(5/5)	100%	(5/5)
12#14	40%	(2/5)	100%	(5/5)	80%	(4/5)
12#15	40%	(2/5)	100%	(5/5)	100%	(5/5)
L1 children: 13-year-olds						
13#1	0%	(0/4)	100%	(4/4)	---	---
13#2	0%	(0/4)	100%	(4/4)	---	---
13#3	25%	(1/4)	100%	(4/4)	---	---
13#4	0%	(0/4)	100%	(4/4)	---	---
13#5	75%	(3/4)	100%	(4/4)	---	---
13#6	75%	(3/4)	100%	(4/4)	---	---
13#7	75%	(3/4)	100%	(4/4)	---	---
13#8	100%	(4/4)	100%	(4/4)	---	---
13#9	100%	(4/4)	100%	(4/4)	---	---
13#10	100%	(4/4)	100%	(4/4)	---	---
13#11	100%	(4/4)	100%	(4/4)	---	---
13#12	100%	(4/4)	100%	(4/4)	---	---
13#13	75%	(3/4)	100%	(4/4)	---	---
13#14	75%	(3/4)	50%	(2/4)	---	---
13#15	100%	(4/4)	100%	(4/4)	---	---

Table B (cont'd). L1 children (comprehension): Individual results on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Negation (target: yes)		*Twee keer*			
			Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
	%	n	%	n	%	n
L1 children: 13-year-olds (cont'd)						
13#16	100%	(4/4)	100%	(4/4)	---	---
13#17	75%	(3/4)	100%	(4/4)	---	---
13#18	75%	(3/4)	100%	(4/4)	---	---
13#19	100%	(4/4)	100%	(4/4)	---	---
13#20	100%	(4/4)	100%	(4/4)	---	---
13#21	100%	(4/4)	100%	(4/4)	---	---
13#22	100%	(4/4)	100%	(4/4)	---	---
13#23	75%	(3/4)	100%	(4/4)	---	---
13#24	100%	(4/4)	100%	(4/4)	---	---
13#25	100%	(4/4)	100%	(4/4)	---	---

Table C. L1 children and adults (comprehension): Distribution of individual response patterns on negation task

Subjects		Average rate of targetlike responses (i.e. yes) within group		No. of individuals with given response pattern		
Age group	**n**	**%**	**SD**	**Rejection**	**Mixed**	**Acceptance**
7-year-olds	14	10.0	25.7	12	1	1
8-year-olds	15	20.0	37.0	11	2	2
9-year-olds	15	65.3	41.7	4	2	9
11-year-olds	15	57.3	44.0	6	0	9
12-year-olds	14	75.4	29.6	1	3	10
13-year-olds	25	77.0	33.8	4	0	21
Adults	14	92.9	14.9	0	2	12

Table D. L1 children children and adults (comprehension): Distribution of individual response patterns on *twee keer* task (scrambling condition)

Subjects		Average rate of targetlike responses (i.e. no) within group		No. of individuals with given response pattern		
Age group	**n**	**%**	**SD**	**Rejection**	**Mixed**	**Acceptance**
7-year-olds	14	25.7	41.8	3	1	10
8-year-olds	16	18.7	38.9	3	0	13
9-year-olds	15	56.0	40.1	6	4	5
11-year-olds	15	60.0	46.0	8	2	5
12-year-olds	14	93.9	22.2	13	0	1
13-year-olds	25	98.0	10.0	24	1	0
Adults	14	90.0	28.0	12	1	1

Table E. L1 children children and adults (comprehension): Distribution of individual response patterns on *twee keer* task (no scrambling condition)

Subjects		Average rate of targetlike responses (i.e. yes) within group		No. of individuals with given response pattern		
Age group	**n**	**%**	**SD**	**Rejection**	**Mixed**	**Acceptance**
7-year-olds	14	81.4	38.0	2	1	11
8-year-olds	16	96.0	15.5	0	1	15
9-year-olds	15	86.7	29.0	1	1	13
11-year-olds	15	88.0	27.0	1	1	13
12-year-olds	14	93.9	17.1	0	1	13
13-year-olds	25	---	---	---	---	---
Adults	14	95.7	11.6	0	1	13

Table F. L2 children (comprehension): Biodata

Subject	Country of origin	Age at time of testing	Age at first exp. (years; months)	Length of exp. (years; months)	Contact with Dutch	Other languages
C1C	USA	10;6,22	7;3	3;3	Mod.	
C2C	USA	7;10,4	7;3	0;7	Ltd.	
C3C	S.Africa	8;4.14	7;2	1;2	Mod.	
C4C	S.Africa	8;9.12	6;6	2;3	Mod.	
C5C[a]	UK	8;2.21	6;4	1;10	Ltd.	
C6C	UK	8;2.21	6;4	1;10	Ltd.	
C7C	UK	11;7.20	5;4	6;3	Ext.	
C8C	UK	7;3.27	6;4	0;11	Mod.	
C9C	UK	7;9.9	4;0	3;9	Ext.	
C10C	Ireland	14;8,29	5	9;8	Mod.	
C11C	UK	10;2.16	6;9	4;5	Mod.	
C12C	Canada	11;2,8	6;2	6;0	Mod.	
C13C	UK	8;10,2	6;2	2;8	Ltd.	
C14C	UK	11;0,3	6;0	5;0	Ext.	
C15C	USA	8;9,18	4;0	4;2	Mod.	
C16C	Australia	7;4,21	6;0	1;4	Ltd.	
C17C	UK	8;5,11	5;10	2;7	Mod.	
C18C	UK	9;10,9	7;1	2;9	Mod.	
C19C	Israel	11;3,8	5	6;3	Ltd.	School French for 1 year
C20C	UK	8;8,21	4	4;4	Ext.	
C21C	UK	10;10,30	6;6	4;0	Mod.	
C22C	Canada	9;3,13	4;4	4;11	Mod.	
C23C	Canada	8;4,27	5;6	2;10	Mod.	
C24C	Australia	10;4,11	6	4;4	Mod.	
C25C	Canada	8;8,26	5;8	3;0	Ext.	
C26C	USA	10;4,5	5;5	4;11	Ext.	School French for 1 year
C27C	UK	7;11,23	6;2	1;9	Mod.	
C28C	UK	9;1,11	6;7	2;6	Mod.	
C29C	UK	11;10,13	5;6[b]	5;4	Mod.	
C30C	Australia	13;3.5	4	9;3	Ext.	
C31C	Finland /USA	13;2.17	5	8;2	Mod.	
C32C	Finland /USA	13;2.17	5	8;2	Mod.	
C33C	UK/USA	14;1,2	5	9;1	Ext.	
C34C	UK	14;11,11	4;6	10;4	Ext.	
C35C	UK	14;8.24	4 ;4	10;4	Ltd.	
C36C	UK	13 ;7.19	4	9 ;7	Mod.	School French for 1 year and Spanish for 2 years

Footnotes for Table F.

[a] Limited (Ltd.) contact with Dutch means that the subject's only contact with Dutch is at school (children) or foreign language classes (adults) and possibly in basic interactions in shops, etc. Moderate (Mod.) contact means that the subject has some Dutch-speaking friends and/or contact with Dutch neighbours. Extensive (Ext.) contact means that subjects meet one or more of the following criteria: for the adults – the subject's partner is a native-speaker of Dutch and interaction with him or her is predominanly in Dutch and/or the subject works in a Dutch-speaking environment (at least some of the time); for the children – the subject has attended a Dutch-speaking school; for either children or adults – the subject lives with one or more native-speakers who speak Dutch to him or her.

[b] Subjects C5C and C6C are twins, as are subjects C31C and C32C.

[c] This subject, who is dyslexic, spent one year in England between arriving in The Netherlands at age 5;6.

Table G. L2 adults (comprehension): Biodata

Subject	Country of origin	Age at time of testing	Age at first exp. (years; months)	Length of exp. (months)	Contact with Dutch	Other languages
A1C	Zimbabwe	10;2,1	9;1	1;1	Ltd.	
A2C	S.Africa / Romania	10;1.11	8;6	1;7	Mod.	
A3C	South Africa	11;2.8	9;10	1;4	Mod.	
A4C	UK	10;5.4	9;6	0;11	Mod.	
A5C	UK	11;1.20	8;7	2;6	Ext.	
A6C	UK	12;11,16	11;10	1;1	Ltd.	
A7C	N.Zealand	13;0,29	12;2	0;10	Mod.	
A8C	Australia	14;4,8	12;5	1;11	Mod.	School German for 1 year?
A9C	USA	14;4,29	11;10	2;6	Ltd.	
A10C	USA	13;10,24	12;4	1;5	Mod.	
A11C	UK	12;8,6	12;4	0;4	Ltd.	
A12C	S.Africa	14;8,2	11;8	2;11	Ltd.	
A13C	N.Zealand	40;10,4	26	14;0	Ext.	
A14C	UK	44;0,3	37	6;10	Ltd.	
A15C	Canada	34;1,5	29;7	4;6	Ext.	
A16C	UK	44;8,10	43	1	Ltd.	Beginner's French
A17C	UK	26;1,1	23;7	2;6	Ext.	Beginner's French, Japanese, Georgian
A18C	UK	22;2.6	21;10	0;4	Mod.	School German and French
A19C	Canada	40;4,2	35;0	5;3	Ext.	Intermediate French, Italian, German and Spanish
A20C	UK	40;3	24	16	Ext.	
A21C	UK	29;3,3	24;5	4;10	Mod.	School German
A22C	UK	29;5.24	24;6	4;11	Mod.	
A23C	UK	29;5,15	23;3	6;2	Ext.	
A24C	Ireland	38;11,4	33	5	Ext.	Advanced German, Intermediate French, Spanish and Gaelic
A25C	USA	30;11.10	27;9	3;2	Mod.	Spanish, 1 year Finnish
A26C	UK	26;4	23;2	3;2	Ext.	School French
A27C	USA	50;0,6	25	25	Ext.	
A28C	UK	39;7.16	26	13	Ext.	German A-level
A29C	USA	36;7,13	34;10	1;9	Mod.	Intermediate Spanish
A30C	UK	47;7,6	34	13	Ext.	School French

Table G (cont'd). L2 adults (comprehension): Biodata

Subject	Country of origin	Age at time of testing	Age at first exp. (years; months)	Length of exp. (months)	Contact with Dutch	Other languages
A31C	UK	25;4	22;11	2;5	Ext.	School French
A32C	UK	10;10,0	8;5	2;5	Mod.	
A33C	USA	10;3,0	8;10	1;5	Mod.	
A34C	UK	9;8,11	8	1;8	Mod.	
A35C[a]	USA	12;0,19	11;7	0;3	Ext.	
A36C[a]	USA	10;6,8	10;3	0;3	Ext.	
A37C	S.Africa	13;8.19	8;6	5;2	Mod.	

[a] These two subjects are brother and sister. Their mother is a native-speaker of Dutch but while they were growing up in the USA, she only spoke to them in English. Two years before the stated age of arrival, both children spent one month in The Netherlands with their Dutch-speaking aunt.

Table H. L2 children (comprehension): Individual results (organised by proficiency) on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Prof. score	Age at time of testing	Negation (target: yes)		*Twee keer*			
					Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
			%	n	%	n	%	n
L2 children: LOW proficiency								
C2C	---	7;10,4	0%	(0/5)	0%	(0/5)	100%	(5/5)
C5C	---	8;2.21	80%	(4/5)	0%	(0/5)	80%	(4/5)
C8C	-1.98	7;3.27	0%	(0/5)	0%	(0/5)	100%	(5/5)
C3C	-1.52	8;4.14	0%	(0/5)	0%	(0/5)	100%	(5/5)
C27C	-1.45	7;11,23	no true negs		0%	(0/5)	100%	(5/5)
C16C	-1.40	7;4,21	no true negs		0%	(0/5)	100%	(5/5)
C35C	-1.36	14;8.24	0%	(0/5)	0%	(0/5)	100%	(5/5)
C4C	-1.25	8;9.12	0%	(0/5)	0%	(0/5)	100%	(5/5)
C13C	-1.19	8;10,2	0%	(0/5)	0%	(0/5)	100%	(5/5)
C23C	-1.11	8;4,27	0%	(0/5)	0%	(0/5)	100%	(5/5)
C17C	-1.07	8;5,11	100%	(5/5)	0%	(0/5)	100%	(5/5)
C6C	-0.95	8;2.21	100%	(5/5)	0%	(0/5)	100%	(5/5)
C11C	-0.66	10;2.16	100%	(5/5)	0%	(0/5)	100%	(5/5)
C18C	-0.51	9;10,9	0%	(0/5)	100%	(5/5)	0%	(0/5)
L2 children: MID proficiency								
C15C	-0.20	8;9,18	100%	(5/5)	0%	(0/5)	100%	(5/5)
C29C	-0.15	11;10,13	40%	(2/5)	100%	(5/5)	100%	(5/5)
C28C	-0.08	9;1,11	40%	(2/5)	60%	(3/5)	80%	(4/5)
C24C	-0.03	10;4,11	0%	(0/5)	60%	(3/5)	100%	(5/5)
C19C	0.17	11;3,8	100%	(5/5)	0%	(0/5)	100%	(5/5)
C36C	0.20	13 ;7.19	20%	(1/5)	100%	(5/5)	60%	(3/5)

Table H (cont'd). L2 children (comprehension): Individual results (organised by proficiency) on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Prof. score	Age at time of testing	Negation (target: yes)		*Twee keer*			
					Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
			%	n	%	n	%	n
L2 proficiency: HIGH proficiency								
C20C	0.60	8;8,21	0%	(0/5)	0%	(0/5)	100%	(5/5)
C7C	0.60	11;7.20	0%	(0/5)	60%	(3/5)	40%	(2/5)
C21C	0.63	10;10,30	100%	(5/5)	100%	(5/5)	0%	(0/5)
C22C	0.70	9;3,13	0%	(0/5)	100%	(5/5)	100%	(5/5)
C32C	0.72	13;2.17	40%	(2/5)	100%	(5/5)	80%	(4/5)
C10C	0.72	14;8,29	0%	(0/5)	0%	(0/5)	100%	(5/5)
C9C	0.74	7;9.9	0%	(0/5)	0%	(0/5)	100%	(5/5)
C14C	0.75	11;0,3	100%	(5/5)	0%	(0/5)	100%	(5/5)
C31C	0.84	13;2.17	0%	(0/5)	80%	(4/5)	100%	(5/5)
C30C	1.02	13;3.5	0%	(0/5)	80%	(4/5)	100%	(5/5)
C33C	1.09	14;1,2	100%	(5/5)	80%	(4/5)	80%	(4/5)
C12C	1.13	11;2,8	0%	(0/5)	0%	(0/5)	100%	(5/5)
C25C	1.41	8;8,26	0%	(0/5)	0%	(0/5)	100%	(5/5)
C1C	1.42	10;6,22	0%	(0/5)	0%	(0/5)	100%	(5/5)
C26C	1.49	10;4,5	40%	(2/5)	100%	(5/5)	0%	(0/5)
C34C	1.71	14;11,11	40%	(2/5)	100%	(5/5)	100%	(5/5)

Table I. L2 adults (comprehension): Individual results (organised by proficiency) on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Prof.score	Age at time of testing	Negation (target: yes)		*Twee keer*			
					Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
			%	n	%	n	%	n
L2 adults: LOW proficiency								
A6C	---	12;11,16	0%	(0/5)	0%	(0/5)	100%	(5/5)
A7C	---	13;0,29	100%	(5/5)	40%	(2/5)	100%	(5/5)
A8C	---	14;4,8	100%	(5/5)	60%	(3/5)	100%	(5/5)
A9C	---	14;4,29	0%	(0/5)	40%	(2/5)	100%	(5/5)
A11C	---	12;8,6	0%	(0/5)	0%	(0/5)	100%	(5/5)
A31C	---	25;4	0%	(0/5)	0%	(0/5)	100%	(5/5)
A10C	-1.89	13;10,24	80%	(4/5)	0%	(0/5)	100%	(5/5)
A16C	-1.49	44;8,10	100%	(5/5)	0%	(0/5)	100%	(5/5)
A3C	-1.31	11;2.8	0%	(0/5)	0%	(0/5)	100%	(5/5)
A18C	-1.26	22;2.6	40%	(2/5)	100%	(5/5)	100%	(5/5)
A34C	-1.25	9;8,11	0%	(0/5)	20%	(1/5)	100%	(5/5)
A4C	-1.20	10;5.4	0%	(0/5)	0%	(0/5)	100%	(5/5)
A32C	-0.90	10;10,0	100%	(5/5)	100%	(5/5)	80%	(4/5)
A25C	-0.85	30;11.10	60%	(3/5)	60%	(3/5)	100%	(5/5)
A29C	-0.84	36;7,13	60%	(3/5)	20%	(1/5)	100%	(5/5)
A1C	-.064	10;2,1	0%	(0/5)	0%	(0/5)	100%	(5/5)
A33C	-0.63	10;3,0	0%	(0/5)	20%	(1/5)	100%	(5/5)
A21C	-0.55	29;3,3	0%	(0/5)	100%	(5/5)	80%	(4/5)
L2 adults: MID proficiency								
A19C	-0.46	40;4,2	0%	(0/5)	100%	(5/5)	60%	(3/5)
A17C	-0.07	26;1,1	0%	(0/5)	0%	(0/5)	100%	(5/5)
A14C	-0.04	44;0,3	0%	(0/5)	0%	(0/5)	100%	(5/5)
A5C	0.11	11;1.20	0%	(0/5)	0%	(0/5)	100%	(5/5)
A22C	0.18	29;5.24	20%	(1/5)	0%	(0/5)	100%	(5/5)
A12C	0.28	14;8,2	0%	(0/5)	0%	(0/5)	100%	(5/5)
A2C	0.46	10;1.11	0%	(0/5)	80%	(4/5)	0%	(0/5)
A15C	0.50	34;1,5	0%	(0/5)	20%	(1/5)	80%	(4/5)

Table I (cont'd). L2 adults (comprehension): Individual results (organised by proficiency) on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Prof.score	Age at time of testing	Negation (target: yes)		*Twee keer*			
					Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
			%	n	%	n	%	n
L2 adults: HIGH proficiency								
A27C	0.62	50;0,6	100%	(5/5)	60%	(3/5)	100%	(5/5)
A13C	0.67	40;10,4	0%	(0/5)	60%	(3/5)	100%	(5/5)
A30C	0.75	47;7,6	40%	(2/5)	100%	(5/5)	100%	(5/5)
A26C	0.84	26;4	20%	(1/5)	100%	(5/5)	100%	(5/5)
A37C	0.87	13;8.19	0%	(0/5)	80%	(4/5)	100%	(5/5)
A36C	0.95	10;6,8	100%	(5/5)	0%	(0/5)	100%	(5/5)
A20C	0.99	40;3	0%	(0/5)	100%	(5/5)	0%	(0/5)
A23C	1.07	29;5,15	80%	(4/5)	0%	(0/5)	100%	(5/5)
A24C	1.16	38;11,4	100%	(5/5)	100%	(5/5)	100%	(5/5)
A28C	1.37	39;7.16	0%	(0/5)	100%	(5/5)	100%	(5/5)
A35C	1.54	12;0,19	0%	(0/5)	0%	(0/5)	100%	(5/5)

Table J. L2 children and adults (comprehension): Distribution of individual response patterns (organised by proficiency) on negation task

Group	Average rate of targetlike (i.e. yes) responses within group		No. of individuals with given response pattern		
	%	SD	Rejection	Mixed	Acceptance
L2 children					
Low	35.0	46.0	7	1	4
Mid	50.0	41.5	2	2	2
High	26.3	39.8	10	3	3
L2 adults					
Low	35.6	43.7	10	3	5
Mid	2.5	7.1	8	0	0
High	40.0	45.6	6	1	4
Native adults	92.9	14.9	0	2	12

Table K. L2 children and adults (comprehension): Distribution of individual response patterns (organised by proficiency) on *twee keer* task (scrambling condition)

Group	Average rate of targetlike (i.e. no) responses within group		No. of individuals with given response pattern		
	%	SD	Rejection	Mixed	Acceptance
L2 children					
Low	7.1	26.7	1	0	13
Mid	53.3	45.0	2	2	2
High	50.0	46.8	8	1	7
L2 adults					
Low	31.1	37.7	3	4	11
Mid	25.0	41.1	2	0	6
High	63.6	43.7	6	2	3
Native adults	90.0	28.0	12	1	1

Table L. L2 children and adults (comprehension): Distribution of individual response patterns (organised by proficiency) on *twee keer* task (no scrambling condition)

Group	**Average rate of targetlike (i.e. yes) responses within group**		**No. of individuals with given response pattern**		
	%	**SD**	**Rejection**	**Mixed**	**Acceptance**
L2 children					
Low	91.4	26.8	1	0	13
Mid	90.0	16.7	0	1	5
High	75.0	36.2	2	1	13
L2 adults					
Low	31.1	6.5	0	0	18
Mid	25.0	35.5	1	1	6
High	63.6	30.2	1	0	10
Native adults	95.7	11.6	0	1	13

Table M. L2 children (comprehension): Individual results (organised by age at time of testing and then by proficiency) on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Prof. score	Age at time of testing	Negation (target: yes)		*Twee keer*			
					Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
			%	n	%	n	%	n
L2 children: 7-year-olds								
C2C	---	7;10,4	0%	(0/5)	0%	(0/5)	100%	(5/5)
C8C	-1.98	7;3.27	0%	(0/5)	0%	(0/5)	100%	(5/5)
C27C	-1.45	7;11,23	no true negs		0%	(0/5)	100%	(5/5)
C16C	-1.40	7;4,21	no true negs		0%	(0/5)	100%	(5/5)
C9C	0.74	7;9.9	0%	(0/5)	0%	(0/5)	100%	(5/5)
L2 children: 8-year-olds								
C5C	---	8;2.21	80%	(4/5)	0%	(0/5)	80%	(4/5)
C3C	-1.52	8;4.14	0%	(0/5)	0%	(0/5)	100%	(5/5)
C4C	-1.25	8;9.12	0%	(0/5)	0%	(0/5)	100%	(5/5)
C13C	-1.19	8;10,2	0%	(0/5)	0%	(0/5)	100%	(5/5)
C23C	-1.11	8;4,27	0%	(0/5)	0%	(0/5)	100%	(5/5)
C17C	-1.07	8;5,11	100%	(5/5)	0%	(0/5)	100%	(5/5)
C6C	-0.95	8;2.21	100%	(5/5)	0%	(0/5)	100%	(5/5)
C15C	-0.20	8;9,18	100%	(5/5)	0%	(0/5)	100%	(5/5)
C20C	0.60	8;8,21	0%	(0/5)	0%	(0/5)	100%	(5/5)
C25C	1.41	8;8,26	0%	(0/5)	0%	(0/5)	100%	(5/5)
L2 children: 9-year-olds								
C18C	-0.51	9;10,9	0%	(0/5)	100%	(5/5)	0%	(0/5)
C28C	-0.08	9;1,11	40%	(2/5)	60%	(3/5)	80%	(4/5)
C22C	0.70	9;3,13	0%	(0/5)	100%	(5/5)	100%	(5/5)
L2 children: 10-year-olds								
C11C	-0.66	10;2.16	100%	(5/5)	0%	(0/5)	100%	(5/5)
C24C	-0.03	10;4,11	0%	(0/5)	60%	(3/5)	100%	(5/5)
C21C	0.63	10;10,30	100%	(5/5)	100%	(5/5)	0%	(0/5)
C1C	1.42	10;6,22	0%	(0/5)	0%	(0/5)	100%	(5/5)
C26C	1.49	10;4,5	40%	(2/5)	100%	(5/5)	0%	(0/5)
L2 children: 11-year-olds								
C19C	0.17	11;3,8	100%	(5/5)	0%	(0/5)	100%	(5/5)
C7C	0.60	11;7.20	0%	(0/5)	60%	(3/5)	40%	(2/5)
C29C	-0.15	11;10,13	40%	(2/5)	100%	(5/5)	100%	(5/5)
C14C	0.75	11;0,3	100%	(5/5)	0%	(0/5)	100%	(5/5)
C12C	1.13	11;2,8	0%	(0/5)	0%	(0/5)	100%	(5/5)

Table M (cont'd). L2 children (comprehension): Individual results (organised by age at time of testing and then by proficiency) on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Prof. score	Age at time of testing	Negation (target: yes)		*Twee keer*			
					Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
			%	n	%	n	%	n
L2 children: 13/14-year-olds								
C35C	-1.36	14;8.24	0%	(0/5)	0%	(0/5)	100%	(5/5)
C36C	0.20	13 ;7.19	20%	(1/5)	100%	(5/5)	60%	(3/5)
C32C	0.72	13;2.17	40%	(2/5)	100%	(5/5)	80%	(4/5)
C10C	0.72	14;8,29	0%	(0/5)	0%	(0/5)	100%	(5/5)
C31C	0.84	13;2.17	0%	(0/5)	80%	(4/5)	100%	(5/5)
C30C	1.02	13;3.5	0%	(0/5)	80%	(4/5)	100%	(5/5)
C33C	1.09	14;1,2	100%	(5/5)	80%	(4/5)	80%	(4/5)
C34C	1.71	14;11,11	40%	(2/5)	100%	(5/5)	100%	(5/5)

Table N. L2 children (comprehension): Distribution of individual response patterns (organised by age at time of testing) on negation task

Subjects		Average rate of targetlike (i.e. yes) responses within group		No. of individuals with given response pattern		
Age group	n	%	SD	Rejection	Mixed	Acceptance
7-year-olds	3	0	0	3	0	0
8-year-olds	10	38.0	49.4	6	0	4
9-year-olds	3	13.3	23.1	2	1	0
10-year-olds	5	48.0	50.2	2	1	2
11-year-olds	5	48.0	50.2	2	1	2
13/14-year-olds	8	30.0	33.8	5	2	1

Table O. L2 children (comprehension): Distribution of individual response patterns (organised by age at time of testing) on *twee keer* task (scrambling condition)

Subjects		Average rate of targetlike (i.e. no) responses within group		No. of individuals with given response pattern		
Age group	n	%	SD	Rejection	Mixed	Acceptance
7-year-olds	3	0	0	0	0	5
8-year-olds	10	0	0	0	0	10
9-year-olds	3	86.7	23.1	2	1	0
10-year-olds	5	52.0	50.2	2	1	2
11-year-olds	5	32.0	46.0	1	1	3
13/14-year-olds	8	67.5	42.7	6	0	2

Table P. L2 children (comprehension): Distribution of individual response patterns (organised by age at time of testing) on *twee keer* task (no scrambling condition)

Subjects		Average rate of targetlike (i.e. yes) responses within group		No. of individuals with given response pattern		
Age group	n	%	SD	Rejection	Mixed	Acceptance
7-year-olds	3	100.0	0	0	0	5
8-year-olds	10	98.0	6.3	1	0	9
9-year-olds	3	60.0	52.9	1	0	2
10-year-olds	5	75.0	50.0	2	0	3
11-year-olds	5	88.0	26.8	0	1	4
13/14-year-olds	8	90.0	15.1	0	1	7

Table Q. L2 adults (comprehension): Individual results (organised by age at time of testing and then by proficiency) on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Prof.score	Age at time of testing	Negation (target: yes)		*Twee keer*			
					Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
			%	n	%	n	%	n
Younger L2 adults: LOW proficiency								
A11C	---	12;8,6	0%	(0/5)	0%	(0/5)	100%	(5/5)
A6C	---	12;11,16	0%	(0/5)	0%	(0/5)	100%	(5/5)
A7C	---	13;0,29	100%	(5/5)	40%	(2/5)	100%	(5/5)
A8C	---	14;4,8	100%	(5/5)	60%	(3/5)	100%	(5/5)
A9C	---	14;4,29	0%	(0/5)	40%	(2/5)	100%	(5/5)
A10C	-1.89	13;10,24	80%	(4/5)	0%	(0/5)	100%	(5/5)
A3C	-1.31	11;2.8	0%	(0/5)	0%	(0/5)	100%	(5/5)
A34C	-1.25	9;8,11	0%	(0/5)	20%	(1/5)	100%	(5/5)
A4C	-1.20	10;5.4	0%	(0/5)	0%	(0/5)	100%	(5/5)
A32C	-0.90	10;10,0	100%	(5/5)	100%	(5/5)	80%	(4/5)
A1C	-.064	10;2,1	0%	(0/5)	0%	(0/5)	100%	(5/5)
A33C	-0.63	10;3,0	0%	(0/5)	20%	(1/5)	100%	(5/5)
Younger L2 adults: MID proficiency								
A5C	0.11	11;1.20	0%	(0/5)	0%	(0/5)	100%	(5/5)
A12C	0.28	14;8,2	0%	(0/5)	0%	(0/5)	100%	(5/5)
A2C	0.46	10;1.11	0%	(0/5)	80%	(4/5)	0%	(0/5)
Younger L2 adults: HIGH proficiency								
A37C	0.87	13;8.19	0%	(0/5)	80%	(4/5)	100%	(5/5)
A36C	0.95	10;6,8	100%	(5/5)	0%	(0/5)	100%	(5/5)
A35C	1.54	12;0,19	0%	(0/5)	0%	(0/5)	100%	(5/5)

Table Q (cont'd). L2 adults (comprehension): Individual results (organised by age at time of testing and then by proficiency) on truth value judgement tasks. Percentage and number of targetlike responses

Subj.	Prof.score	Age at time of testing	Negation (target: yes)		*Twee keer*			
					Scrambling condition (target: no)		Non-scrambling condition (target: yes)	
			%	n	%	n	%	n
Older L2 adults: LOW proficiency								
A31C	---	25;4	0%	(0/5)	0%	(0/5)	100%	(5/5)
A16C	-1.49	44;8,10	100%	(5/5)	0%	(0/5)	100%	(5/5)
A18C	-1.26	22;2.6	40%	(2/5)	100%	(5/5)	100%	(5/5)
A25C	-0.85	30;11.10	60%	(3/5)	60%	(3/5)	100%	(5/5)
A29C	-0.84	36;7,13	60%	(3/5)	20%	(1/5)	100%	(5/5)
A21C	-0.55	29;3,3	0%	(0/5)	100%	(5/5)	80%	(4/5)
Older L2 adults: MID proficiency								
A19C	-0.46	40;4,2	0%	(0/5)	100%	(5/5)	60%	(3/5)
A17C	-0.07	26;1,1	0%	(0/5)	0%	(0/5)	100%	(5/5)
A14C	-0.04	44;0,3	0%	(0/5)	0%	(0/5)	100%	(5/5)
A22C	0.18	29;5.24	20%	(1/5)	0%	(0/5)	100%	(5/5)
A15C	0.50	34;1,5	0%	(0/5)	20%	(1/5)	80%	(4/5)
Older L2 adults: HIGH proficiency								
A27C	0.62	50;0,6	100%	(5/5)	60%	(3/5)	100%	(5/5)
A13C	0.67	40;10,4	0%	(0/5)	60%	(3/5)	100%	(5/5)
A30C	0.75	47;7,6	40%	(2/5)	100%	(5/5)	100%	(5/5)
A26C	0.84	26;4	20%	(1/5)	100%	(5/5)	100%	(5/5)
A20C	0.99	40;3	0%	(0/5)	100%	(5/5)	0%	(0/5)
A23C	1.07	29;5,15	80%	(4/5)	0%	(0/5)	100%	(5/5)
A24C	1.16	38;11,4	100%	(5/5)	100%	(5/5)	100%	(5/5)
A28C	1.37	39;7.16	0%	(0/5)	100%	(5/5)	100%	(5/5)

Act-out task: Experimental items

t1	*rollen 'roll'*	
Context:	four different-coloured marbles, marble run with three different tubes for marbles to go down	
Scrambling	*Je moet een knikker twee keer rollen* You should roll a (certain) marble twice	[1-object]
Context:	four different-sized coins and cardboard box with three holes to roll the coins into	
No scrambling	*Je moet twee keer een muntje rollen* You should roll a(ny) coin twice	[2-object]
t2	*omdraaien 'turn'*	
Context:	four transparent plastic jars filled with oil and glitter	
Scrambling	*Je moet een potje twee keer omdraaien* You should turn a (certain) jar twice	[1-object]
Context:	four different-coloured egg-timers	
No scrambling	*Je moet twee keer een zandloper omdraaien* You should turn an(y) egg-timer twice	[2-object]
t3	*wegblazen 'blow away'*	
Context:	four different-coloured cotton wool balls, two straws, one cardboard box from which top and back have been removed. Three "gates" have been cut out from the front	
Scrambling	*Je moet een watje twee keer wegblazen* You should blow away a (certain) cotton wool ball twice	[1-object]
Context:	four different-coloured balls, two straws, one cardboard box from which top and back have been removed. Three "gates" have been cut out from the front	
No scrambling	*Je moet twee keer een balletje wegblazen* You should blow away a(ny) ball twice	[2-object]
t4	*indrukken 'press'*	
Context:	four different plastic cars in holders with buttons, which release the car when pressed; the action includes returning the car to its holder (so that the same object could be used twice)]	
Scrambling	*Je moet een knopje twee keer indrukken* You should press a (certain) button twice	[1-object]
Context:	four plastic bicycle horns, in different animal shapes, which, when pressed, make a squeaking noise	
No scrambling	*Je moet twee keer een diertje indrukken* You should press a(ny) bicycle horn twice	[2-object]
t5	*gooien 'throw'*	
Context:	four different-coloured rings, three target sticks	
Scrambling	*Je moet een ring twee keer gooien* You should throw a (certain) ring twice	[1-object]
Context:	four different-coloured balls with container	
No scrambling	*Je moet twee keer een bal gooien* You should throw a(ny) ball twice	[2-object]

Act-out: Fillers

As in original

f1 two white beans and three black ones, toy teapot
Je moet de witte boontjes in het potje doen [2-object]
You should put the white beans in the teapot

f2 seven coloured Duplo blocks, two of which are green
Je moet van de groene blokjes één toren maken [1-object]
You should make one tower with the green blocks

f3 two windmills, one with a yellow middle and one with a green middle
Je moet tegen het molentje met het gele middenstuk blazen [1-object]
You should blow against the windmill with the green middle

f4 six different-coloured marbles, two of which are blue, four toy cups
Je moet de blauwe knikkers in een kopje doen [2-object]
You should put the blue marbles in the cup

f5 four different-coloured giant beads, two of which are blue
Je moet de blauwe kralen aan elkaar vast maken [2-object]
You should string the blue beads

f6 four different-coloured spoons, one plastic cup
Je moet een lepel in de kop leggen [1-object]
You should put a spoon in the cup

Fillers where two objects are moved (rather than two actions being carried out)

f7 four different-looking cows, a pen
Je moet twee koeien het weiland in laten lopen [2-object]
You should make two cows go into the field

f8 four different-looking horses, a fence
Je moet twee paarden over het hek laten springen [2-object]
You should make two horses jump over the fence

f9 four different-looking birds, a tree
Je moet twee vogels over de boom laten vliegen [2-object]
You should make two birds fly over the tree

f10 four different-looking monkies, a car
Je moet twee apen in de auto laten stappen [2-object]
You should make two monkeys get in the car

f11 four different-looking cars, tower made of lego bricks
Je moet twee auto's rond het torentje laten rijden [2-object]
You should make two cars drive round the tower

Fillers were action is carried out three times

f12	four different-looking tigers, a sheep	
Scrambling	*Je moet het schaap drie keer over een tijger laten springen*	[1-object]
No scrambling	*Je moet drie keer het schaap over een tijger laten springen*	[1-object]
	You should make the sheep jump over a tiger three times	
f13	four different-looking rabbits, a girl	
Scrambling	*Je moet het meisje drie keer een konijn laten zoenen*	[1-object]
No scrambling	*Je moet drie keer het meisje een konijn laten zoenen*	[1-object]
	You should make the girl kiss a rabit three times	
f14	four different-looking lego-bricks, a man	
Scrambling	*Je moet de man drie keer een blokje laten aanraken*	[1-object]
No scrambling	*Je moet drie keer de man een blokje laten aanraken*	[1-object]
	You should make the man touch a lego-brick three times	
f15	four different-looking toy plates, a bird	
Scrambling	*Je moet het vogeltje drie keer om een bordje laten lopen*	[1-object]
No scrambling	*Je moet drie keer het vogeltje om een bordje laten lopen*	[1-object]
	You should make the bird walk around the plate three times	
f16	four different-looking birds, a horse	
Scrambling	*Je moet het paard drie keer over een vogel laten springen*	[1-object]
No scrambling	*Je moet drie keer het paard over een vogel laten springen*	[1-object]
	You should make the horse jump over a bird three times	

Table R. L1 adults (comprehension): Individual results on act-out task. Percentage and number of 1-object responses

Subj.	Age at time of testing	Scrambling condition (target: 1-object)		Non-scrambling condition (target preference: 2-object)	
		%	n	%	n
N#1	23	100%	(5/5)	0%	(5/5)
N#13	20	100%	(5/5)	0%	(0/5)
N#14	18	40%	(2/5)	0%	(0/5)
N#15	22	80%	(4/5)	0%	(0/5)
N#16	23	80%	(4/5)	0%	(0/5)
N#17	20	100%	(5/5)	0%	(0/5)
N#18	20	100%	(5/5)	0%	(0/5)
N#19	22	40%	(2/5)	0%	(0/5)
N#20	22	100%	(5/5)	60%	(3/5)
N#21	21	100%	(5/5)	0%	(0/5)
N#22	24	100%	(5/5)	0%	(0/5)
N#23	32	100%	(5/5)	0%	(0/5)
N#24	21	100%	(5/5)	0%	(0/5)
N#25	26	100%	(5/5)	40%	(2/5)

Table S. L2 adults (comprehension): Individual results (organised by proficiency) on act-out task. Percentage and number of 1-object responses

Subj.	Prof.score	Age at time of testing	Scrambling condition (target: 1-object response)		Non-scrambling condition (target preference: 2-object response)	
			%	n	%	n
L2 adults: LOW proficiency						
A31C	---	25;4	100%	(5/5)	40%	(2/5)
A18C	-1.26	22;2.6	20%	(1/5)	0%	(0/5)
A25C	-0.85	30;11.10	100%	(5/5)	100%	(5/5)
A29C	-0.84	36;7,13	100%	(5/5)	100%	(5/5)
A21C	-0.55	29;3,3	100%	(5/5)	80%	(4/5)
L2 adults: MID proficiency						
A19C	-0.46	40;4,2	100%	(5/5)	80%	(4/5)
A22C	0.18	29;5.24	20%	(1/5)	40%	(2/5)
L2 adults: HIGH proficiency						
A27C	0.62	50;0,6	80%	(4/5)	20%	(1/5)
A30C	0.75	47;7,6	100%	(5/5)	100%	(5/5)
A26C	0.84	26;4	100%	(5/5)	40%	(2/5)
A20C	0.99	40;3	100%	(5/5)	20%	(1/5)
A23C	1.07	29;5,15	100%	(5/5)	100%	(5/5)
A24C	1.16	38;11,4	100%	(5/5)	60%	(3/5)
A28C	1.37	39;7.16	80%	(4/5)	0%	(0/5)

Table T. L2 adults (comprehension): Distribution of individual response patterns (organised by proficiency) on act-out task. Scrambling condition.

Group	1-object responses for group	No. of individuals with given response pattern		
	%	1-object	2-object	Mixed
Low	84.0	4	1	0
Mid	60.0	1	1	0
High	84.3	7	0	0
Controls	88.6	12	0	2

Table U. L2 adults (comprehension): Distribution of individual response patterns (organised by proficiency) on act-out task. No scrambling condition.

Group	2-object responses for group	No. of individuals with given response pattern		
	%	1-object	2-object	Mixed
Low	64.0	3	1	1
Mid	60.0	1	0	1
High	49.6	2	3	2
Controls	14.2	1	11	2

SAMENVATTING IN HET NEDERLANDS

Dit proefschrift vergelijkt drie verschillende groepen van leerders van het Nederlands: vroege tweede taalverwervers (T2ers), late T2ers en T1-kinderen. Deze groepen worden bestudeerd met betrekking tot hun verwerving van een bepaalde eigenschap van de doeltaal, namelijk scrambling van het lijdend onderwerp. De linguïstische ontwikkeling van de drie soorten verwervers wordt vergeleken om de vraag te beantwoorden of vroege T2-, late T2- en T1-verwerving in wezen hetzelfde proces (of dezelfde processen) bevat(ten).

De kwestie of late T2-verwerving door hetzelfde aangeboren mechanisme als T1-verwerving, d.w.z. door Universele Grammatica (UG), wordt gestuurd heeft de afgelopen twintig jaar centraal gestaan binnen het onderzoeksgebied van (generatieve) T2-verwerving. Bij veel studies die deze vraag behandelen is er sprake van een directe vergelijking van late T2-verwerving, d.w.z. T2-verwerving door volwassenen, met T1-verwerving door kinderen. In de middenpunt van deze dissertatie staat echter de vroege T2-verwerver, hier gedefinieerd als 'een leerder wiens eerste contact met de T2 tussen de leeftijd van vier en seven plaatsvindt'. In dit boek worden vroege T2ers enerzijds met late T2ers en anderzijds met T1-kinderen vergeleken. De basis voor deze vergelijkingen wordt in **Hoofdstuk 1** gelegd, waarin de grondgedachte achter de hier uiteengezette studie, een voorstel dat zijn oorsprong in Schwartz (1992) vindt, wordt besproken. Schwartz stelt dat vroege T2-verwervingsonderzoek een belangrijke bijdrage kan leveren aan het debat over de rol van UG versus de rol van algemene leerprincipes in late T2-verwerving. Haar voorstel is dat aangenomen dat (i) vroege T2-verwerving gedreven wordt door UG en dat, gegeven de voor kinderen beschikbare UG, (ii) algemene leermethodes relevanter zijn voor late T2ers, dan (iii) een vergelijking tussen de ontwikkelingsstadia van vroege T2ers en die van late T2ers met dezelfde eerste taal bewijs kan leveren voor dan wel tegen de betrokkenheid van UG in de volwassen verwerving van een tweede taal.

Verder wordt in het eerste hoofdstuk een gedetailleerd en kritisch overzicht gegeven van de beschikbare literatuur over vroege T2-verwerving. Wat het duidelijkst uit dit overzicht naar boven komt is dat de initiële toestand van het T2 kind gekarakteriseerd wordt door invloed uit de eerste taal. Er bestaan slechts vier studies die vroege en late T2ers direct met elkaar vergelijken. De resultaten van deze studies verschillen wat betreft de ontwikkelingsstadia die T2-kinderen en volwassen doorlopen. De bespreking in Hoofdstuk 1 vestigt de aandacht op meerdere methodologische kwesties en abstraheert hieruit een lijst van zes desiderata voor studies van vroege T2, late T2 en (vroege) T1 zoals de huidige.

Hoofdstuk 2 introduceert de eigenschap van de doeltaal die hier centraal staat, namelijk het zg. *scramblen* van het lijdend voorwerp in het

Nederlands. De verplaatsing van een indefiniete lijdend voorwerp-NP (object) dient ertoe betekenis te disambigueren: gescramblede indefiniete objecten zoals in (1)-a krijgen een specifieke interpretatie, terwijl niet-gescramblede indefiniete objecten onspecifiek geïnterpreteerd worden, zoals in (1)-b.

(1) a. Simon heeft [een roos] twee keer besproeid
b. Simon heeft twee keer [een roos] besproeid

Ook het scramblen van definiete object-NPs over negatie geeft betekeniseffecten: negatieve zinnen met een gescramblede definiet worden over het algemeen geïnterpreteerd als uitdrukkingen van zinsnegatie, zoals in (2)-a, terwijl met een niet-gescramblede definiet er een interpretatie ontstaan met contrastieve negatie, zoals in (2)-b.

(2) a. Bob heeft [het onkruid] niet weggegooid
b. Bob heeft niet [het onkruid] weggegooid (maar de jonge tomatenplanten)

In Hoofdstuk 2 worden verschillende syntactische, *discourse*-gerelateerde en semantische benaderingen van scrambling besproken. In dit proefschrift wordt geprobeerd zo neutraal mogelijk te blijven wat betreft de analyse van scrambling. Gebaseerd op ideeën van Neeleman and Reinhart (1998), worden de scramblingeigenschappen van definiete NPs in negatieve zinnen geanalyseerd als volgend uit de wisselwerking tussen negatie en focus. Wat betreft de analyse van indefiniete objecten, gescramblede indefinieten wordt een ander semantisch type toegekend dan niet-gescramblede indefinieten, namelijk dat van een 'vrije variabele' volgens de aanpak van Van Geenhoven (1998) of dat van kwantificationele NPs volgens de aanpak van de Hoop (1992). Vanuit een syntactisch oogpunt wordt scrambling gezien als beweging naar een positie buiten de VP.

In **Hoofdstuk 3** wordt beargumenteerd dat de verwerving van de interpretatie van gescramblede indefiniete objecten een zg. *poverty-of-the-stimulus* probleem presenteert voor Engelssprekende T2ers: de kennis dat gescramblede indefiniete objecten uitsluitend specifiek geïnterpreteerd kunnen worden is niet aanwezig in de T2-input, kan niet worden overgedragen uit de eerste taal (waar scrambling niet in voorkomt), en is niet het onderwerp van instructie in de klas. Hetzelfde kan worden gesteld (tot op zekere hoogte tenminste) voor de verwerving van gescramblede definiete NPs in negatieve zinnen. Hieruit volgt dat indien de T2-volwassenen en -kinderen dezelfde kennis van interpretatieve condities op scrambling als native speakers laten zien, hun *interlanguage*-grammatica's op eenzelfde wijze gereguleerd zijn als L1-verweving.

In Hoofdstuk 3 worden verder eerdere studies van de verwerving van scrambling besproken. De beschikbare data uit die studies suggereren dat T1-kinderen en T2-volwassenen een stadium van 'geen scrambling' en/of 'optioneel scrambling' doorlopen voordat ze zoals native speakers scramblen en dat de productie van gescramblede voorwerpen vóór het begrip hiervan komt (Krämer 2000; Schaeffer 2000). Krämer (2000) beargumenteert dat voor de doel-interpretatie van gescramblede indefiniete objecten, discourse-integratie nodig is; zo is de afwezigheid van T1-gelijkend begrip bij kinderen het gevolg van een gelimiteerde mogelijkheid tot discourse-integratie. De studie in deze dissertatie bouwt voort op dit eerdere onderzoek door (i) systematisch de productie en het begrip van scrambling, met al de bijbehorende interpretatieve effecten, bij T2-volwassenen en –kinderen te onderzoeken en door (ii) nieuwe T1-data te leveren die enkele van de methodologische problemen met eerdere studies rechtzet en het onderzoeksdomein vergroot naar oudere kinderen.

De volgende voorspellingen worden geformuleerd, gebaseerd op Schwartz (1992), de bevindingen van de eerdere verwervingsstudies en het poverty-of-the-stimulus probleem dat de condities op scrambling met zich mee brengen:

VOORSPELLING ❶. VROEGE T2 ~ LATE T2: ONTWIKKELINGSSTADIA

T2-kinderen en T2-volwassenen zullen dezelfde ontwikkelingsstadia doorlopen.

VOORSPELLING ❷. VROEGE T2 ~ LATE T2: DISCOURSE/PRAGMATISCHE FACTOREN

a. T2-kinderen binnen dezelfde leeftijdsgroep als T1-kinderen met gelimiteerde discourse-integratie zullen van T2-volwassenen afwijken wat betreft hun interpretatie van gescramblede indefiniete objecten.

b. T2-kinderen buiten deze leeftijdsgroep zullen geen verschil met T2-volwassenen laten zien.

VOORSPELLING ❸. VROEGE T2 ~ VROEGE T1: ONTWIKKELINGSSTADIA

Als gevolg van de invloed van T1 zullen T2-kinderen andere ontwikkelingsstadia doorlopen als T1-kinderen.

VOORSPELLING ❹. VROEGE T2 ~ VROEGE T1: DISCOURSE / PRAGMATISCHE FACTOREN

T2-kinderen binnen dezelfde leeftijdsgroep als T1-kinderen met gelimiteerde discourse-integratie zullen eenzelfde soort gedrag m.b.t. de interpretatie van gescramblede indefiniete objecten vertonen als T1-kinderen

Deze voorspellingen zijn getest met een productie-experiment en twee taken waarin de proefpersonen waarheidswaarde-oordelen moeten geven. Hierbij is steeds dezelfde methodologie gebruikt bij alledrie de groepen leerders. De data van de beoordelingstaak zijn bovendien aangevuld met een taak voor een gedeelte van de T2-volwassenen, waarin de proefpersonen beschreven opdrachten moeten uitvoeren (een zg. *act-out* taak). Voor indefiniete objecten wordt met het productie experiment – waarover in **Hoofdstuk 5** wordt bericht

– onderzocht of leerders de doelvorm (gescrambled versus niet-gescrambled) in de relevante context (specifiek versus niet-specifiek) produceren. Voor definiete objecten bepaalt dit experiment of een gescramblede vorm geproduceerd wordt in een context waar zinsnegatie bedoeld is. De begripstaak – waarover in **Hoofdstuk 6** wordt bericht – vraagt of leerders de doel-interpretatie (specifiek versus niet-specifiek) aan de relevante vorm (gescrambled versus niet-gescrambled) toekennen. De eerste beoordelingstaak met waarheidswaardes test de interpretatie die de leerder toekent aan indefiniete objecten die over negatie gescrambled zijn, zoals in (3). De tweede beoordelingstaak en de act-out taak testen het scramblen over de bijwoordelijke bepaling van frequentie *twee keer*, zoals in (4) geïllustreerd wordt.

(3) De jongen heeft een vis niet gevangen

(4) a. Het meisje heeft een aap twee keer gekieteld
b. Het meisje heeft twee keer en aap gekieteld

De T2-proefpersonen hebben een extra taak uitgevoerd om hun taalvaardigheid Nederlands te bepalen. De taalvaardigheidsmaat, die in **Hoofdstuk 4** wordt gepresenteerd, bestaat uit drie sub-scores, namelijk verbale dichtheid (het gemiddelde aantal van finiete en niet-finiete werkwoorden per T-eenheid), lexicale diversiteit (Guiraud's maat, i.e. $V/\sqrt{N}$) en de relatieve hoeveelheid aan foutvrije uitingen (het aantal foutvrije uitingen gedeeld door het aantal T-eenheden). Deze scores worden gecombineerd om tot een enkele taalvaardigheidsscore voor elke proefpersoon te komen.

De belangrijkste bevindingen zijn als volgt. In productie blijken T2-kinderen en T2-volwassenen dezelfde ontwikkelingsstadia te doorlopen, hetgeen voorspelling ❶ bevestigt. De stadia voor definiete en specifieke indefiniete objecten worden gegeven in (5) en de stadia voor niet-specifieke indefinieten zijn te vinden in (6).

(5) **De ontwikkelingsstadia voor de verwerving van gescramblede definiete en specifieke indefiniete objecten in het Nederlands door Engelssprekende T2-kinderen en T2-volwassenen**

Stadium	Beschrijving	Geproduceerde woordvolgordes
I	Geen scrambling	uitsluitend Negatie-Werkwoord-Object
II-a	Geen scrambling	Negatie-Object-Werkwoord *en* Negatie-Werkwoord-Object
II-b	Geen scrambling	Uitsluitend Negatie-Object-Werkwoord
III-a	Optioneel scrambling	Object-Negatie-Werkwoord *en* Negatie-Object-Werkwoord
III-b	Scrambling	Uitsluitend Object-Negatie-Werkwoord

(6) **Ontwikkelingsstadia voor de verwerving van niet-gescramblede niet-specifieke indefiniete objecten in het Nederlands door Engelssprekende T2-kinderen en T2-volwassenen**

Stadium	Beschrijving	Geproduceerde woordvolgordes
I	Geen scrambling	uitsluitend Negatie-Werkwoord-*een* NP
II	Doelvolgorde zonder gebruik van *geen*-vorm	uitsluitend negatie-*een* NP-werkwoord
III-a	Optioneel gebruik van *geen*-vorm	*geen* NP-Werkwoord *en* Negatie-*een* NP-werkwoord
III-b	Gebruik van doelvolgorde en *geen*-vorm	*geen* NP-werkwoord

Kinderen en volwassenen met mid-hoge en hoge taalvaardigheid scramblen definiete en specifieke indefinieten soms of altijd in de relevante contexten; de proefpersonen met hoge taalvaardigheid scramblen over het algemeen altijd.

De verzameling data van vijf jaar oude T1-kinderen geeft aan dat ook zij het belangrijke onderscheid tussen specifieke en niet-specifieke indefinieten kunnen maken. Oftewel, ze scramblen de eerstgenoemde, terwijl ze de laatgenoemde in basispositie laten staan. De ontwikkelingsstadia voor de T1-kinderen worden hieronder gegeven: voor definiete en specifieke indefiniete objecten in (7) en voor niet-specifieke indefinieten in (8).

(7) **Ontwikkelingsstadia voor de verwerving van gescramblede definiete en specifieke indefiniete objecten in het Nederlands door T1-kinderen.**

Stadium	Beschrijving	Geproduceerde woordvolgordes
I	Geen scrambling	Uitsluitend Negatie-Object-Werkwoord
II-a	Optioneel scrambling	Object-Negatie-Werkwoord *en* Negatie-Object-Werkwoord
II-b	Scrambling	Uitsluitend Object-Negatie-Werkwoord

(8) **Ontwikkelingsstadia voor de verwerving van niet-gescramblede niet-specifieke indefiniete objecten in het Nederlands door T1-kinderen.**

Stadium	Beschrijving	Geproduceerde woordvolgordes
I	Doelvolgorde zonder gebruik van *geen*-vorm	uitsluitend Negatie-*een N*-Werkwoord
II-a	Optioneel gebruik van *geen*-vorm	*geen N*-Werkwoord *en* Negatie-*een N*-Werkwoord
II-b	Gebruik van doelvolgorde en *geen*-vorm	uitsluitend *geen N*-Werkwoord

Het enige verschil tussen de T1- en T2-stadia is de aanwezigheid van een Negatie-Werkwoord-Object stadium in de T2-sequentie. Dit is, zo wordt in

Hoofdstuk 3 beargumenteerd, het resultaat van de invloed van T1. Voorspelling ❸ is zo bevestigd.

Wat betreft begrip blijkt dat T1-kinderen tot wel 13 jaar een niet-specifieke interpretatie aan gescramblede indefiniete objecten toekennen in de taak met negatie, en dat ze deze van de doel-interpretatie afwijkende betekenis tot hun 11e jaar bij de *twee keer*-taak laten zien. Vanaf het negende levensjaar verschillen de kinderen uit de T1-groep echter niet significant van de moedertaal-volwassenen. Bij de *twee keer*-taak suggereren de vroege T2-data eenzelfde soort leeftijdsgerelateerde ontwikkeling als bij de T1-kinderen: de (op het moment van testen) jongere kinderen verschillen significant van de oudere kinderen en de moedertalige proefpersonen, en anders dan de T2-volwassenen met hoge taalvaardigheid verschillen de T2-kinderen met hoge taalvaardigheid significant van de T1-volwassenen. Bij de act-out taak, die ook scrambling over *twee keer* betrof, laten de T2-volwassenen met hoge taalvaardigheid dezelfde voorkeuren zien als de moedertaal-volwassenen. Dat wil zeggen dat ze significant minder vaak een enkel object manipuleren in de niet-scrambling-conditie dan in de scrambling-conditie. Deze beide verzamelingen resultaten staan in contrast met de resultaten van de beoordelingstaak met scrambling over negatie. In deze taak zijn de responsen van de T2-kinderen en T2-volwassenen consistent slecht. Er is geen ontwikkeling waar te nemen, noch naarmate de leeftijd stijgt bij de vroege T2ers noch naarmate taalvaardigheid stijgt bij de beide T2 groepen. Dit betekent dat de voorspellingen ❷ en ❹ bevestigt worden voor de *twee keer*-data, maar niet voor de negatie-data. De relatief slechte uitvoering van de negatie taak door de T2ers kan worden toegerekend aan een combinatie van het mogelijk ongepaste gebruik van negatie in het experimentele scenario (cf. Gualmini 2003) en de invloed van T1.

In **Hoofdstuk 7**, worden de productie- en begripsresultaten voor gescramblede indefinieten met elkaar vergeleken. De discrepantie tussen productie en begrip die in eerdere studies is opgemerkt wordt op groepsniveau bevestigd voor zowel de T1- als de T2-kinderen en op het niveau van individuen voor die T2-kinderen die aan beide experimenten deelgenomen hebben. Voor deze 12 proefpersonen kan de volgende generalisatie gemaakt worden: proefpersonen die (soms of altijd) T1-achtig begrip laten zien, laten (soms of altijd) T1-achtige productie zien. Het omgekeerde geldt overigens niet: proefpersonen die (soms of altijd) T1-achtige productie laten zien, laten niet (altijd of soms) T1-achtig begrip zien. In Hoofdstuk 7 worden verschillende mogelijke verklaringen hiervoor onder de loep genomen, waaronder methodologische zaken als de hoge mate waarin de aandacht wordt gevestigd op sommige van de objecten in het experimentele scenario en het ongepaste gebruik van negatie. Een andere mogelijke verklaring die bekeken wordt is de gelimiteerde mogelijkheid bij de proefpersoon tot discourse-integratie, zoals Krämer (2000) en Miller and Schmitt (2004) al bespreken.

De observatie dat er Engelssprekende T2-kinderen en -volwassen zijn die consistent de niet-specifieke lezing voor gescramblede indefiniete objecten afwijzen laat zien dat deze leerders erin geslaagd zijn de poverty-of-the-stimulus te overwinnen. Op zijn beurt suggereert dit dat T2-ontwikkeling gebruik maakt van dezelfde mechanismen als T1-ontwikkeling. Uiteindelijk wordt in Hoofdstuk 7 de volgende these voorgeslagen: leerders maken gebruik van een *blocking* principe (Williams 1997) om de niet-specifieke lezing uit te sluiten voor het gescramblede indefiniete object. Het blocking principe garandeert dat vanaf het moment dat leerders weten dat objecten in het Nederlands de mogelijkheid hebben tot beweging naar een gescramblede positie, de niet-specifieke lezing geblokkeerd zal zijn voor de gescramblede indefiniet. Deze 'blokkering' is uitsluitend het resultaat van het feit dat er voor deze lezing een 'goedkopere' optie bestaat, namelijk de niet-gescramblede volgorde.

CURRICULUM VITAE

Sharon Unsworth was born on the 10th of October, 1974, in Preston, Lancashire, in the UK. Before starting university, Sharon spent six months teaching English in Lézignan-Corbières, France. From 1994 to 1998, she attended the University of Newcastle upon Tyne, where she obtained a first class joint honours BA degree in French and German. Whilst at university, Sharon taught French to English-speaking children at a local primary school. As part of her studies, she also spent a year teaching English in Mönchengladbach, Germany. She completed the Certificate of English Language Teaching to Adults (CELTA) in 1998. From 1998 to 2001, Sharon was a graduate student at the University of Durham. It was there that she obtained an MA in Language Acquisition (with distinction) and where she started her PhD. In 2001, Sharon joined the Utrecht Institute of Linguistics, initially as a Marie Curie fellow and subsequently as a PhD researcher. This thesis is a result of the work she carried out there.