Dynamic Development in Speaking Versus Writing in Identical Twins

Chan, HuiPing; Verspoor, Marjolijn; Vahtrick, Louisa

Published in:
Language Learning

DOI:
10.1111/lang.12107

Publication date:
2015

LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

Dynamic Development in Speaking Versus Writing in Identical Twins

HuiPing Chan,[a] Marjolijn Verspoor,[a,b] and Louisa Vahtrick[a]

[a]University of Groningen and [b]University of the Free State, South Africa

Taking a dynamic usage-based perspective, this longitudinal case study compares the development of sentence complexity in speaking versus writing in two beginner Taiwanese learners of English (identical twins) in an extensive corpus consisting of 100 oral and 100 written texts of approximately 200 words produced by each twin over 8 months. Three syntactic complexity measures were calculated: mean length of T-unit, dependent clauses per T-unit, and coordinate phrases per T-unit. The working hypothesis was that (a) the learners' oral texts would become more complex sooner than their written texts and that (b) the two learners would show similar developmental patterns. We found that these two learners initially demonstrated syntactic complexity in their oral language rather than in their written language, yet over time they were found to exhibit inverse trends of development. This observation was confirmed with dynamic modeling by means of a hidden Markov model, which allowed us to detect moments of self-organization in the learners' spoken and written output (i.e., moments where the interaction among various measures changes and takes on a new configuration).

Keywords writing; speaking; syntactic complexity; developmental patterns; HMM; usage-based; self-organization

Introduction

Researchers working within a dynamic usage-based perspective on language learning and use view second language (L2) development as a dynamic process because all factors involved, including the amount of meaningful exposure and motivation, affect the process continuously (e.g., Larsen-Freeman, 2006; de Bot, Lowie, & Verspoor, 2007; Verspoor, de Bot, & Lowie, 2011). Within this

---

The elicitation instruments used for the study can be accessed by readers in the IRIS digital repository (http://www.iris-database.org).

Correspondence regarding this article should be addressed to: HuiPing Chan, University of Groningen, Department of English, P.O. Box 716, Groningen, 9700 AS, Netherlands. E-mail: huipingchan1981@gmail.com

DOI: 10.1111/lang.12107

perspective, language itself is viewed as a complex usage-based system with many subsystems, all of which are interconnected and may demonstrate shifts in the developmental trajectory at different stages. Such shifts can be described as the inevitable reorganization a dynamic system undergoes throughout its lifetime. In previous L2 studies that have employed a dynamic usage-based perspective, the focus has been on the variability found within individual measures and how these measures may interact over time (Spoelman & Verspoor, 2010; Verspoor, Lowie, & van Dijk, 2008), on the variation among learners (Larsen-Freeman, 2006), or on developmental peaks that may be demonstrative of overuse of targeted structures (van Dijk, Verspoor, & Lowie, 2011). The current article focuses on stages of development, punctuated by moments of reorganization. We investigate at what points in the learning process the interaction among different linguistic subsystems shifts.

Different subsystems may show various interrelationships over time: precursor, competitive, and supportive (Caspi, 2010; Van Geert, 2008). A precursor relationship occurs when a particular subsystem needs to be in place before another can begin to develop. For example, for children learning their first language (L1), one-word utterances (i.e., individual words) precede the occurrence of two- or three-word utterances. A competitive relation occurs when one subsystem develops at the expense of another; for example, in L1 development, a growth in multiword utterances is usually accompanied by a dip in the rate of learning new single words (Robinson & Mervis, 1998). A supportive relationship occurs when subsystems grow in tandem. For example, once lexical and syntactic subsystems have matured and become automated, they no longer compete and therefore develop synchronously. In other words, the interaction between different sub-systems may change over time, in what is called moments of self-organization (Caspi, 2010; Spoelman & Verspoor, 2010; Van Geert, 2008; Verspoor et al., 2008; Verspoor, Schmid, & Xu, 2012).

Previous research examining the development of different aspects of L2 learner language has revealed that languages are likely to develop lexically before they develop syntactically at various stages of the learning process, with this development proceeding in wave-like patterns; only after one measure has shown a rise and/or fall, the other one will. A cross-sectional study conducted by Verspoor et al. (2012) found that the beginner (Common European Framework of Reference levels A1.1 to A1.2) and intermediate (levels B1.1 to B1.2) learners in their study developed more in lexicon than in grammar, while the low-intermediate learners (levels A1.2 to A2) demonstrated more syntactic than lexical development. Similarly, in a longitudinal study with more advanced students, Caspi (2010) showed that three of her four learners exhibited comparable patterns in the learning of lexicon and syntax, which developed in wave-like patterns with lexical development preceding syntactic development. Two points can be taken from these studies. The first is that, at least when investigating L2 learners of English, we should assume wave-like patterns for different subsystems, and the second point is that it is possible to distinguish shifts in wave-like patterns among the subsystems at specific times, indicating new stages in each learner's developmental trajectory.

A usage-based perspective does not necessarily suggest that there is a predetermined path of development, but it assumes that each learner will have to discover his or her own path through trial and error. In other words, development and change are individually "owned." This developmental trajectory is inevitably accompanied with variability in the learning of a language skill as any subsystem that has not fully matured is likely to be affected when learners' attention is drawn to other subsystems, and learners will differ in the amount of variability they demonstrate as they develop various language subsystems (van Dijk et al., 2011). In a longitudinal study following four Chinese learners of English, Larsen-Freeman (2006) found that although the averages for measures of vocabulary complexity (adjusted type-token ratio) and grammatical complexity (average number of clauses per T-unit) demonstrated upward trends, no two learners were alike and all showed a great deal of variability. Moreover, Verspoor et al. (2012) found that there is more variability (intralearner changeability over time) and variation (interlearner differences) among beginners than among more advanced learners.

Therefore, to be able to investigate development in this manner, we must use extensive individual-based longitudinal data, as any data averaging or data reduction across multiple learners will conceal the interindividual variability as well as the interactions between each learner's linguistic subsystems. In the current study, we compare identical twins, who live in the same home and have been attending the same classes in school. Twin studies normally compare monozygotic (MZ, or identical) twin pairs with dizygotic (DZ, or fraternal) twin pairs in order to investigate the effect that genetic factors have on language (Segal, 2010; Stromswold, 2006). The current study is not a traditional twin study in this sense, in that only one pair of MZ twins was examined. The majority of twin studies have found identical twins to perform more similarly than fraternal twins in linguistic development (e.g., Stromswold, 2006). In stating that our participants are identical twins, we are not invoking the much-maligned equal environments assumption (Plomin, Defries, McClern, & McGuffin, 2008), which argues that MZ and DZ twins both share equal environments, so any similarities in developmental patterns found in MZ twins must be due to shared genetics. Instead, we merely hypothesize that twins who share 100% of their genes and who have been raised in a similar environment are more likely than any other pair of learners to exhibit similar developmental patterns (Hayiou-Thomas, 2008). In this article, we use several syntactic complexity measures to investigate twin learners' development in both speaking and writing to determine (a) whether syntactic complexity develops in speaking earlier than in writing and (b) whether twin learners develop syntactic complexity in their speech and writing in a similar manner.

The current study differs from previous ones that deal with dynamic development over time in that it did not focus on the variability in individual measures, nor on the specific interactions between these measures. Instead, this study focused on moments of self-organization, that is, the moments in the learning process where the various linguistic subsystems reorganize themselves and where the resulting variability marks a new stage in a developmental trajectory.

Background

Complexity in Speaking and Writing in L1 and L2

Most research comparing the development of complexity in speaking and writing has targeted learners' lexicon rather than syntax, but as research findings are often explained in terms of cost of production, we assume they may also apply to syntax. Bourdin and Fayol (1994, 1996, 2002) conducted several studies with both children and adults. For example, using word-recall tasks, they showed that children from the second and fourth grades recalled fewer words in writing than in speaking, which was attributed to relatively little practice children had in handwriting and spelling (Bourdin & Fayol, 1994). Bourdin and Fayol (1996) then investigated the recall of sentences produced in speaking and writing tasks and found similar results, namely, children performed less accurately in writing than in speaking. In order to move beyond the level of merely recalling oral and written language, Bourdin and Fayol (2002) then conducted a follow-up study that tested more complex production processes (such as sentence construction, lexical choice, and expression of meaning) in two conditions. Educated L1 adults were presented with sets of words (some semantically linked, others semantically unrelated) and were asked to produce sentences with these words in speaking and writing. Participants' recall of the presented words was similar in speaking and writing tasks, and the tasks did not differ in number of grammatically correct sentences. However, the results from speaking and writing differed in the conceptual domain, such that there was more elaboration of ideas and more construction of a coherent framework between the semantically unrelated words in the oral than in the written mode. Bourdin and Fayol (2002) therefore concluded that written production is more "costly" than oral production, even in adults. Interpreting their results in the framework of capacity theories in production (Fayol, 1999; McCutchen, 1996, 2000), which is in line with our own dynamic view of limited resources and possible competition for attention, these researchers suggest that writing texts demands a certain degree of attention. Put differently, writing texts with weakly associated elements demands an extra processing capacity to maintain and organize the information in working memory and/or an ability to adopt a strategic approach to the composition of texts. This additional processing burden, according to this view, must be added to the cost of production.

Conversely, the situation may be quite different in the L2. Learners may develop their speaking and writing skills at the same time, or in some cases written language may be acquired before spoken language. For example, Milton and Hopkins (2006) found that L2 learners of English have a smaller oral vocabulary size ($M = 2260$), compared to their written vocabulary size ($M = 2655$). They also found that oral and written vocabulary size varied with English proficiency, such that learners of lower L2 proficiency levels tended to have better phonological than orthographic vocabulary knowledge, whilst learners of higher proficiency levels demonstrated more orthographic than phonological vocabulary knowledge. However, the development of spoken and written vocabulary is likely related to the nature of input L2 learners receive. Hudelson (1984) points out that instructed learners, in particular, who are first exposed more to the written than to the oral language, may be able to read and write but may not be fully proficient in speaking.

In the field of L2 development, there are few studies that have compared the development of oral and written language. This may be because these skills, although they rely on the same linguistic resources (Levelt, 1989), are in fact very different, both in terms of how they are produced and perceived (e.g., sequences of sounds vs. letters). In L2 research, many complexity indices of speech and writing, both lexical and syntactic, have been widely investigated and defined (Bulté & Housen, 2012; Ortega, 2003; Polio, 1997; Quinn & Nation, 2007; Wolfe-Quintero, Inagaki, & Kim, 1998). The few studies that have compared L2 development in oral and written language using several complexity measures have yielded findings that are for the most part in line with those of Bourdin and Fayol (1994, 1996, 2002). Yu (2010) used the D index, defined as a Type/Token ratio corrected by means of a curve-fitting procedure that alleviates problems with text length (Malvern & Richards, 2002), to compare lexical diversity of written texts and interviews from 25 advanced learners of English. Contrary to what was expected, written language, which allows learners more chances to plan and organize their production, was not associated with greater lexical diversity, compared to oral language. Chan, Lowie, and de Bot (2014) subsequently compared two lexical measures in the oral and written language of the same pair of identical twins (beginner learners of English) as in the current study. The two measures were the D index of lexical diversity and the V_size measure of lexical difficulty (which counts the number of words in different frequency bands and generates an estimated value of difficulty of word use). In contrast with Yu's findings, Chan et al. found that the twin learners did demonstrate lower lexical diversity (D index), but not lower lexical difficulty (V_size), in their oral language when compared to their written language. Chan et al. concluded that the difference in the linguistic performance between oral and written language was most likely due to the extra time allowed for the writing task. In other words, the learners had more time to think during writing about which words they could use and were therefore more likely to use a greater variety of words.

In syntactic measures, Dykstra-Pruim (2003) observed university learners of German in three tasks: an oral narrative task, a written narrative task, and a written grammar task. Over three semesters, she compared different elements of grammar in both oral and written output. She found that the average number of attempted clauses of the more difficult type (with inversion after a preposed element or verb position at the end of a subordinate clause) per subject was higher in oral than in written language; however, in the written mode these clauses were significantly more accurate. Because our study does not deal with accuracy, it is important to note that, in Dykstra-Pruim's data, oral language was syntactically more complex than written language. In contrast, both Yuan and Ellis (2005), who used passive phrases as an index, and Robinson (2007), who used infinitival phrases as an index, found that writing is associated with higher syntactic complexity than speaking and attributed this finding to the extra time allowed in writing.

The fact that some studies (Dyksta-Pruim, 2003; Yu, 2010) found oral language to be equally complex or more complex than written language is rather surprising given that learners have more time in producing written than oral language to think about the manner and form in which they wish to express their thoughts. One study that shows that time-on-task may indeed be a relevant factor in language production is by Yuan and Ellis (2003). Focusing on oral language, they analyzed the effect of planning on complexity. Most relevant for the current study are Yuan and Ellis's results in the planning conditions (with or without online planning). In the online planning condition, there was no time limit allowing for online monitoring. In the no-online planning condition, there was a time limit, which forced the speakers to speed up their production. Yuan and Ellis found that the speakers who had no time limit produced oral language with greater syntactic complexity than the speakers who had a time limit. The authors concluded that the time allowed for online planning has a positive effect on the syntactic complexity of oral production.

Although we might expect greater complexity in writing than in speaking because of the extra time for planning (Biber, Gray, & Poonpon, 2011; Weigle, 2002), oral texts may not necessarily be less complex, as the nature of written and oral complexity is different. Spoken language may actually contain more dependent clauses (which are taken as a sign of complexity in the current study) than written language, which in turn contains more elaborate noun phrases. However, Biber et al. (2011) looked specifically at academic written English, which is of a much higher proficiency level than what our participants would be likely to use. As Byrnes, Maxim, and Norris (2010) point out, the choice of appropriate complexity measures will be related to the proficiency level of the learners, which will result in researchers using different measures for different levels of L2 proficiency.

Only a few studies have thus far found more complexity in written than oral production (Robinson, 2007; Yuan & Ellis, 2005), but these studies have focused on isolated measures. In the remaining L1 and L2 studies discussed previously, oral language has been found to be generally more complex than written language, despite the extra planning time learners often have when producing written language. Therefore, we assume that, when looking at more general, holistic complexity measures, such as mean length of T-unit and amount of subordination, there will be a difference in favor of oral rather than written production. We also assume that if speaking and writing tasks are similar in topic and in the amount of time allowed (no time pressure), oral language production will be more complex. Although online planning or editing is possible in both types of production, written production may be more demanding of cognitive resources (Bourdin & Fayol, 2002). In order to test these assumptions, the current study investigated how L2 oral and written language develops in syntactic complexity over time (8 months) and whether the two twin learners show similar developmental paths.

Complexity Measures in Language Development

Complexity—which, along with accuracy and fluency, is one of several indices used to document language development (see Hakuta, 1976; Larsen-Freeman, 1976; Ortega, 2003; Wolfe-Quintero et al., 1998)—has been extensively investigated in several domains. For example, Bulté and Housen (2012) presented a taxonomic model of L2 complexity that includes three components: propositional complexity, discourse-interactional complexity, and linguistic complexity. The propositional complexity is the amount of information (number of ideas) present in the production (Ellis & Barkhuizen, 2005); discourse-interactional complexity (which only exists in learners' dialogic discourse) refers to the number and the types of exchanges that learners engage in; and linguistic complexity refers to the degree of output elaboration in breadth and in depth. In our study, we focused on linguistic complexity.

Within the dimensions of linguistic complexity, syntactic complexity has received by far the most attention. Syntactic complexity represents "the range of syntactic forms that surface in language production and the degree of sophistication of such forms" (Ortega, 2003, p. 493). Several cross-sectional studies have investigated how measures of syntactic complexity could potentially distinguish groups of differing proficiency levels in their L2 production (e.g., Bardovi-Harlig & Bofman, 1989; Larsen-Freeman, 1978) while longitudinal studies have examined how to track the development of syntactic complexity in writing over time (e.g., Ishikawa, 1995; Norris & Ortega, 2000), showing that mean length of T-unit, mean length of clause, clauses per T-unit, and number of dependent clauses per clause were better markers of syntactic complexity than other metrics. However, as argued by Norris and Ortega (2009), across 16 studies in task-based research, these measures did not consistently show that they could distinguish adjacent English proficiency levels, due in part to different research designs, different L1s of the participants, and different measurement formats. Therefore, it is difficult to say what metrics should be implemented for what level of English proficiency, in what type of language learning, and for which L1. In line with these findings, Bulté (2013) summarizes how an L2 system may develop from a dynamic usage-based perspective:

The L2 system of a learner can develop (expand, grow) in many different directions (i.e., along many different dimensions and in many different subsystems). Words can be added to the lexicon (more items, more variety), different meanings or functions of words can be learned (more components within an item, and more relations), more specific words for restricted contexts and situations (higher sophistication), independent clause and simple sentence structures and word order, coordination and subordination of clauses (horizontal and hierarchical relationships), subordination within phrases, verb paradigms . . . All of these changes make the L2 system of a learner more complex, and this is (or at least can be) also reflected in a more complex L2 production (p. 100).

To thoroughly represent syntactic complexity in L2 writing, complexity metrics should measure different dimensions of complexity and provide distinctive characteristics with as little overlap as possible to avoid redundancy across measures. Therefore, we should only include metrics that represent independent traits and that do not correlate highly with other metrics. A range of distinctive measures were deployed by Norris and Ortega (2009) with five central foci: length of production units, amount of subordination, amount of coordination, sophistication and acquisitional timing of grammatical forms used in production, and total frequency of use of certain forms considered to be sophisticated. Length of production is generally calculated by dividing words by a production unit (e.g., clause, sentence, T-unit); this measure has been widely used in child language acquisition since Brown (1973). Amount of subordination is computed by dividing the number of instances of a subordinate clause by a production unit, for example, the mean number of dependent clauses per T-unit. Amount of coordination, proposed by Bardovi-Harlig (1992), is a metric computed by dividing the number of coordinate clauses by the total number of combined clauses. However, according to Bulté and Housen (2012), a coordination index so calculated does not result in positive numbers for more coordinate constructions; rather it results in greater numbers when there is more subordination. Therefore, they suggest that coordination should be calculated independently of subordination.

Different metrics also tend to reveal different developmental trajectories at different proficiency levels of learners. However, mean length of T-unit and degree of subordination have been shown to correlate strongly with different levels of proficiency over time, especially at the lower levels; therefore, these two metrics were used as target measures in the current study. For our third dimension of complexity, we sought a measure that would tap into another dimension of complexity and found that the number of coordinate phrases did not significantly correlate with either mean length of T-unit or degree of subordination. We assume that coordinate phrases are rather easy to form and may therefore occur earlier than subordinate clauses. In addition, Bulté (2013) argues that it is important to include both clausal subordination and phrasal complexity measures in order to properly assess L2 complexity, as they are both measures of syntactic complexification that do not occur together, nor develop in parallel. Therefore, the third targeted measure of syntactic complexity in this study was degree of coordination. Thus, by using three different dimensions of syntactic complexity (mean length of T-unit, degree of subordination, degree of coordination), we examined in which mode (oral or written) the language of two beginner learners of English (Common European Framework of Reference

Table 1 Score on the General English Proficiency Test: Gloria and Grace

	Gloria	Grace
Listening (max. 120)	112	112
Speaking (max. 100)	80	80
Reading (max. 120)	108	105
Writing (max. 100)	88	82

levels A2 to B1) demonstrates more complexity and whether these learners (identical twins) show similar developmental patterns in these measures.

The Study

Participants

Gloria and Grace (pseudonyms) are two female identical twins, aged 15 at the time of the study. For 10 years, they had attended school in Taiwan in the same English class with the same English teacher, where English classes were taught in Chinese with a focus on grammar. In other words, until the current study began, they had mainly received written input in English. At the beginning of the study, they had a very similar English proficiency level (see Table 1), as measured by the General English Proficiency Test (Wu, 2012). According to the Big Five Personality Test (http://www.outofservice.com/bigfive), they also had similar personalities. They were both open to new experiences (creative, curious, and original); they were sociable, friendly, and talkative; and they were nervous, high-strung, and worried.

Procedure

The study lasted for 8 months. Throughout the study, the researcher asked the participants to obtain extra aural exposure through media in three stages: low, high, and low. According to learner self-reports in their diaries, they received about 2–5 hours per week of extra input until data point 20; 5–15 hours per week until data point 56; and again 2–5 hours per week until the last data point. Frequent informal Facebook contact with the researcher about the content of the movies that the learners watched confirmed the statements in the self-reports. The amount of exposure was manipulated for another study targeting the effect of exposure on vocabulary development (Chan, Lowie, & de Bot, in press).

During the 8 months, the participants produced oral and written texts approximately three times per week, which was usually on Fridays, Saturdays, and Sundays. For each participant, 100 oral texts and 100 written texts were gathered. The topics, selected from the list of standard TOEFL tests, were of the same genre. All the topics were presented to the two participants at the beginning of the study. Sample topics for speaking and writing are shown in (1) and (2), respectively.

(1) Which of the following statements do you agree with? Some believe that TV programs have a positive influence on modern society. Others, however, think that the influence of TV programs is negative. What TV programs have a positive influence? Why? What TV programs have a negative influence? Why?

(2) Do you agree or disagree with the following statement? With the help of technology, students nowadays can learn more information and learn it more quickly. Use specific reasons and examples to support your answer.

In order to motivate and remind the participants to obtain extra exposure to English and to do the speaking and writing tasks, the researcher created a private Facebook group for the project, which only the researcher as well as the participants and their parents had access to. The researcher reminded the twins every week to record themselves and to write texts. Recordings were sent through e-mail, and written texts were posted in the Facebook account. To keep the participants motivated, the researcher reacted to the content of each text, but no corrective feedback on linguistic form was given for either oral or written texts.

All texts were prepared for processing using Lu's automatic syntactic complexity analyzer (Lu, 2010). The analyzer is designed to investigate the syntactic complexity in writing, with 14 indices of syntactic complexity calculated. Lu provides clear descriptions of sentences, clauses, dependent clauses, T-units, complex T-units, and complex nominals (pp. 481–484). For our study, length of production unit (T-unit), subordination, and coordination were used as measures. These measures represent the three main categories employed extensively in L2 acquisition to index language development (e.g., Lu, 2010; Norris & Ortega, 2009). A T-unit is "one main clause plus any subordinate clause or nonclausal structure that is attached to or embedded in it" (Hunt, 1970, p. 4). T-units also include sentence fragments punctuated by the writer (Bardovi-Harlig & Bofman, 1989; Tapia, 1993). Because it is only possible to parse trees one by one using Tregex (Levy & Galen, 2006), it was specified that a T-unit can only occur within a sentence punctuated by the writer (Homburg, 1984; Ishikawa, 1995). A dependent clause is defined as a finite adjective, adverbial, or nominal clause (Cooper, 1976; Hunt, 1965; Kameen, 1979). Nonfinite verb phrases are excluded in the definition of clauses (e.g., Bardovi-Harlig & Bofman, 1989).

As far as coordinate phrases are concerned, only adjective, adverb, noun, and verb phrases are counted in coordinate phrases (Cooper, 1976).

Of the three selected measures, the mean length of T-unit (MLT) is a general complexity measure while the number of dependent clauses per T-unit (DC/T) is a more specific measure of complexity. These two measures partly overlap because T-units tend to become longer as dependent clauses are added. Therefore, in order to find a third measure that would tap into a different, less-overlapping dimension of complexity, we ran correlations with the remaining 12 syntactic complexity measures from the analyzer output and selected the one with the lowest correlation with the MLT, which was the number of coordinate phrases per T-unit (CP/T). Based on Verspoor et al. (2012), we assumed that coordinate phrases show less development over time and are less indicative of linguistic complexity than the other two measures.

Lu (2010) tested the reliability of the analyzer using essays written by Chinese learners of English at the university level. As these learners were advanced writers, their writings mostly contained errors of word use (e.g., collocations) rather than errors of grammatical completeness. Thus, errors occurring within the boundaries of a structural unit led to little misanalysis by the parser or little misrecognition of the production units. However, as Lu points out, writing samples of beginner learners that contain errors of syntactic completeness should be carefully preprocessed. Therefore, all oral texts (each about 200 words in length) were first transcribed by the researcher. To avoid redundancy in oral production, filled pauses (e.g., *mm* and *er*), dysfluencies (e.g., repetitions, restarts, repairs), and utterances that did not involve linguistic meaning or form (e.g., laughter) were deleted. Then, both the oral and written texts were preprocessed for the analyzer, mainly to enable the analyzer to count the units, meaning that incorrect punctuation and incomplete sentence structures were corrected minimally (see Appendix for examples). All other errors remained to keep the data as original as possible. After this initial processing, the text files were submitted one by one to the analyzer to obtain the values for the targeted syntactic measures.

Analyses

To test our predictions, both traditional and dynamic statistical analyses were used. For overall differences, the oral and written texts were first compared with an independent two-sample *t* test. To identify learning stages, the data were first visually inspected for general patterns. Then, to find the exact moments of reorganization, the dataset was modeled using a hidden Markov model (HMM). As pointed out by van Dijk et al. (2011), visual inspection is first needed to get a feel for the data and to stipulate hypotheses that can later be tested in modeling. Because raw data for the three measures had different numerical ranges (MLT: 7–10; DC/T: .2–3; CP/T: 0–.8), the data were first normalized to the 0–1 scale to be able to observe the relationships between the measures on a common scale. Because a high degree of variability in each measure obscures any general, discernible pattern, a moving average trend line was added for six data points. These trend lines were visually inspected to see if there were shifts in the configuration of the measures. Discernible shifts point to possible boundaries to be found through modeling.

Most longitudinal dynamic studies so far have aimed at investigating the trajectory of separate measures (cf. Larsen-Freeman, 2006; Verspoor et al., 2008, 2012) or the relations between these measures, such as precursor, competitive, or supportive (Caspi & Lowie, 2013; Chan et al., in press). For example, the modeling carried out by Lowie, Caspi, van Geert, and Steenbeek (2011) on longitudinal data involved the following steps: observe the empirical data by means of visual inspection; hypothesize explicit parameters for initial conditions, precursor relationships, and other degrees of interaction among the measures; run simulations; and perform manual adjusting of the parameters until high correlations among the observed and modeled data are found, which can then confirm the hypothesized interaction among the measures. These data analyses have revealed important characteristics of language development, such as individual differences, the meaning of variability, and trade-off effects.

The current study took the dynamic analyses in a somewhat different direction, both in terms of the aims and type of modeling. The aim was to find moments of self-organization, the moments where the interaction among measures shifts and takes on a new configuration, indicating the beginning of a new learning stage. However, as the number of observed measures increases, it becomes more difficult to visually observe any changes in the interaction between the measures and then model them, as pointed out by Lowie et al. (2011). Therefore, we used a dynamic model, an unsupervised HMM, which detects discontinuity patterns based on data. In this model, a string of data (or a value) for all measures is first analyzed to detect patterns of change. After that, the model finds the best stage sequence for the detected changes and identifies the data point where a shift in the complex system occurs, indicating stage boundaries.

There are two types of HMMs, supervised and unsupervised. The specific type of model depends on whether there are clearly identifiable labels for the categories. The supervised HMM typically deals with correctly tagged texts, such as those tagged for parts of speech (PoS), where probabilities of certain sequences of PoS are calculated. For example, when you see the article "the," there is a 40% chance that the next word is a noun. The Constituent Likelihood Automatic Word-tagging System (http://ucrel.lancs.ac.uk/claws/) pioneered the HMM-based tagging for parts of speech (Garside, 1987; Garside & Smith, 1997). In contrast, unsupervised models, which we used in this study, deal with data strings that have no preset identifiable labels, in our case, values expressing differences in degree. In an unsupervised HMM, the values are visible but no knowledge of the probability matrices for initial, transition, and observation data is presupposed in the model. A large number of studies in the field of speech recognition use unsupervised HMMs (e.g., Novotney, Schwartz, & Ma, 2009; Park & Glass, 2008; Zhang & Glass, 2009).

By using HMM, it is possible to evaluate different numbers of learning stages and explore which number of stages reveals changes over time the most accurately; however, the current study a priori postulated three stages in the modeling for two reasons. We visually inspected interactions between measures graphically and observed three to five stages, depending on the mode (speaking or writing). But more importantly, there is a minimum number of data points required for a stage. In other words, the number of data points in each stage must be minimally two to three times the number of observed measures. As we worked with six measures (three spoken, three written), there had to be a minimum of 12 to 18 data points in each stage, and preferably more. Therefore, we decided on three stages.

The modeling software was based on Chan and Lee (2013), and one of the authors programmed the software in Perl for a Linux environment for the current study. There are two major algorithms involved in the process of looking for the highest probability in the data (as discussed below). The Baum-Welch algorithm, also known as the "expectation maximization" (EM) algorithm (Rabiner, 1989), aims to find the most likely changes in learning stages by computing transitional probabilities, emission probabilities, and initial probabilities, that is, the expected stages of development that will occur. The Viterbi algorithm (Ryan & Nudd, 1993) then determines the best stage sequence, based on these combined probabilities.

The six observable measures, the MLT, DC/T, and CP/T at each data point for both writing and speaking, were used to train the model. Only raw data were used for the modeling. The model was initialized with a linear structure where state 1 can transit to state 1 or state 2, state 2 can transit to state 2 or state 3, and state 3 can only transit to state 3. The parameters were adjusted to find the best model with the Baum-Welch algorithm (Rabiner, 1989). The training was discontinued when the model converged, that is, when further

Table 2 Mean of Syntactic Measures in Speaking and Writing: Gloria and Grace

Measure		*Speaking*		*Writing*		
		M	*SD*	*M*	*SD*	*Sig.*
MLT	Gloria	13.2	3.1	10.0	1.2	$p < .01$
	Grace	14.2	3.6	12.9	3.0	$p < .01$
DC/T	Gloria	.9	.4	.5	.2	$p < .01$
	Grace	.9	.4	.9	.4	
CP/T	Gloria	.2	.2	.1	.1	$p < .01$
	Grace	.3	.2	.2	.2	$p < .01$

iteration resulted in no significant change in the model. After obtaining a set of parameters, a single best stage sequence was calculated using the Viterbi algorithm (Ryan & Nudd, 1993). The modeling output yielded three sources of information, namely, (a) data points at the beginning and end of each stage, (b) means for each measure at each stage, and (c) covariances between measures at each stage.

Results

The mean of Gloria's oral language was 199.4 words (*SD* = 34.2) per text; the mean of Gloria's written language was 185.3 words (*SD* = 48.4) per text. The mean of Grace's oral language was 200.6 words (*SD* = 38.0) per text; the mean of Grace's written language was 153.7 words (*SD* = 54.3) per text. Traditional statistical analyses by means of *t* tests (Table 2) indicated that, for Gloria, all three measures of syntactic complexity in speaking were significantly higher than those in writing, but that this was not the case for Grace. In Grace's data, MLT and CP/T in speaking were significantly higher than those in writing, but the DC/T in speaking and writing was almost the same.

To analyze change over time, raw data were plotted in Excel with a moving average trend line of six data points to visually inspect the data. Because it is difficult to see interactions occurring between six measures, we first examined the data in each mode for each learner separately before we combined them for each learner. We added arrows where we observed a qualitatively different configuration of the measures that might suggest a shift. We looked for both changes in the measure over time (clear rise or fall) and changes in the interaction among the measures (where they cross). However, we must keep in mind that the trend lines represent moving averages of six data points, so they cannot specify the exact data point of these changes.

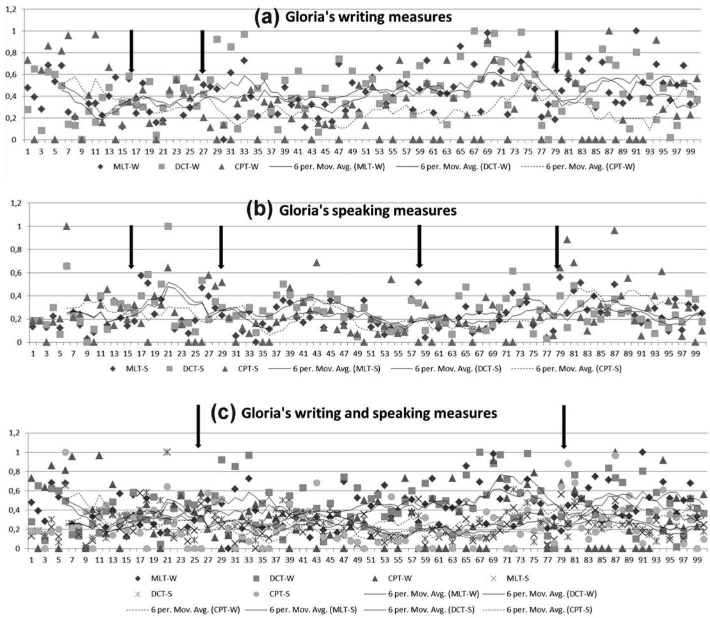

Figure 1 Excel plots of Gloria's normalized MLT, DC/T, and CP/T measures, with moving averages over time and arrows pointing to possible shifts in configuration.

In Gloria's writing (Figure 1a), we observed such shifts around data points 15, 27, and 79, with CP/T decreasing at each of these points. In Gloria's speaking (Figure 1b), we observed such moments around data points 15, 27, 59, and 79, with CP/T decreasing in each case except at data point 79, where it increased. In the combined writing and speaking graph (Figure 1c), it was difficult to identify the specific measures, but shifts in configuration seem to occur around data points 27, 45, and 79.

To quantitatively determine the exact points of the shifts in the configuration of the six measures in Gloria's data, we employed the modeling procedure, as described above (HMM). Table 3 shows the results; the left columns lists the three learning stages, as found by the HMM, and the next columns show the mean for each measure during that learning stage in the speaking and writing conditions. As illustrated in this table, Gloria's shifts seem to have occurred after data points 27 and 78, both of which had also been visually observed in the writing and speaking and the combined measures in the Excel graphs. Note also that Gloria's MLT, DC/T, and CP/T in speaking were consistently higher than those in writing in all three learning stages.

Table 3 Mean values for the three measures of syntactic complexity in speaking and writing in three stages (Gloria)

	M – MLT		*M* – DC/T		*M* – CP/T	
Gloria	Speaking	Writing	Speaking	Writing	Speaking	Writing
Stage 1(datapoints 1–27)	13.9	9.6	1.0	.5	.2	.1
Stage 2(datapoints 28–78)	12.4	10.0	.9	.6	.1	.1
Stage 3(datapoints 79–100)	14.1	10.4	.9	.5	.3	.1

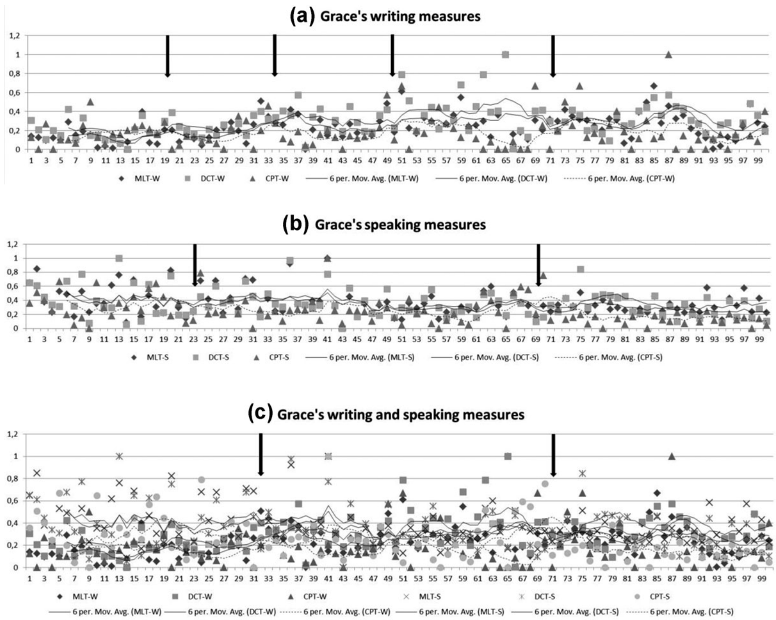

Figure 2 Excel plots of Grace's normalized MLT, DC/T, and CP/T measures, with moving averages over time and arrows pointing to possible shifts in configuration.

In Grace's measures (Figure 2), the shifts in configuration were overall visually less clear than in Gloria's writing. In the writing measures (Figure 2a), we observed shifts around data point 19, where the CP/T crosses and decreases, data point 33, where there is a clear difference in the distance between the trend lines, from converging to diverging, data point 49, where the trend lines start diverging from each other again, and at 71, where they converge again. In Grace's speaking (Figure 2b), we observed such moments around data points

Table 4 Mean of syntactic complexity in speaking and writing in three stages: Grace

	M – MLT		*M* – DC/T		*M* – CP/T	
Grace	Speaking	Writing	Speaking	Writing	Speaking	Writing
Stage 1(1-31)	15.6	11.0	1.0	.6	.4	.2
Stage 2(32-70)	13.7	13.9	.9	1.0	.3	.2
Stage 3(71-100)	13.4	13.5	.8	.9	.2	.2

23, where the trend lines converge, and 69, where the CP/T trend line crosses again and ascends. In the combined writing and speaking graph (Figure 2c), it was difficult to discern shifts in the configuration, but the HMM analysis (see below) pointed to shifts in configuration after data points 31, where the CP/T makes a clear dip below all the other measures, and after 71, where the trend lines seem to converge before they diverge again.

To quantitatively determine the exact points of the shifts in the configuration of the six measures in Grace's data, we again employed modeling (HMM), with the output data shown in Table 4. The left columns again show the three learning stages, as detected by the model, and the next columns present the mean for each measure during each learning stage in the speaking and writing conditions. As Table 4 shows, Grace's largest shifts appeared to occur after data points 31 and 71, neither of which had been found through visual inspection of the data, likely because our visual analysis was based on trend lines that were averaged over six data points. Grace's shifts occurred at different times than Gloria's, suggesting that Grace's second stage of development was shorter than Gloria's. In addition, Grace's MLT and DC/T were higher in speaking than in writing in the first learning stage, but lower in the second and third learning stages. Grace's CP/T was higher in speaking than in writing in the first and second learning stages, but lower in the third one. In Grace's data, there also seemed to be a shift from higher syntactic complexity in speaking to higher syntactic complexity in writing in the third learning stage.

Table 5 shows Spearman correlations (based on normalized covariances obtained from the HMM) for each measure per learning stage for speaking and writing. If correlations between speaking and writing measures are positive, the constructs likely develop synchronously in speaking and writing. If correlations are negative, they develop asynchronously. If there is no meaningful association, the relationship appears to be neutral. As indicated by the trend lines in Figures 1 and 2 and the output of the HMM (Tables 3 and 4), different developmental patterns were found for the twins. Gloria's MLT in speaking and writing

Table 5 Spearman correlation between writing and speaking for each measure of syntactic complexity (Gloria and Grace)

	MLT	DC/T	CP/T
Gloria			
Stage 1 (datapoints 1–27)	.07 ($p = .8$)	−.05 ($p = .7$)	−.10 ($p = .6$)
Stage 2 (datapoints 28–78)	−.07 ($p = .6$)	.09 ($p = .4$)	−.14 ($p = .3$)
Stage 3 (datapoints 79–100)	−.23 ($p = .2$)	.38 ($p = .1$)	.47 ($p = .02$)
Grace			
Stage 1 (datapoints 1–31)	−.09 ($p = .7$)	.18 ($p = .3$)	−.05 ($p = .8$)
Stage 2 (datapoints 32–70)	−.14 ($p = .4$)	−.03 ($p = .9$)	−.24 ($p = .1$)
Stage 3 (datapoints 71–100)	.08 ($p = .6$)	−.03 ($p = 1.0$)	−.08 ($p = .5$)

developed synchronously in the first stage and asynchronously in the last two stages, while Grace's MLT in speaking and writing developed asynchronously in the first two stages but synchronously in the last stage. With respect to DC/T, Gloria's DC/T in speaking and writing developed asynchronously in the first stage but synchronously in the last two stages, whilst Grace's DC/T in speaking and writing developed synchronously in the first stage but asynchronously in the last two stages. And in terms of CP/T, Gloria's CP/T in speaking and writing developed asynchronously in the first stage but synchronously in the last two stages whilst Grace's CP/T in speaking and writing developed asynchronously in all three stages. Therefore, we may conclude that the directions of the correlations for the MLT and DC/T measures, except for MLT in stage 2, in speaking and writing were in opposite directions for the twins.

Discussion

This study explored two questions with respect to the development of syntactic complexity in identical twin learners of L2 English, namely, (a) does syntactic complexity develop first in spoken or written language? and (b) do the two learners develop in similar ways? For the first question, we hypothesized that syntactic complexity would develop in speaking first and then in writing. This hypothesis was based mainly on the findings of previous studies (e.g., Dykstra-Pruim, 2003) and was formulated despite our knowledge that our learners had been exposed more to written than oral language before the study and in contrast to previous research showing that writing allows for more time for online planning, compared to speaking (Yuan & Ellis, 2005). We found support for this hypothesis, with significance testing showing that, in the majority of cases, measures of syntactic complexity were significantly greater in speaking than in writing. We had assumed that the participants had less time for online planning in their speaking, but as our results for syntactic complexity suggest, they had sufficient time to plan online.

Therefore, our findings are in line with other previous research, including the study by Bourdin and Fayol (2002) who demonstrated that L1 children as well as adults tend to elaborate more in oral than in written language, mainly because writing seems to come at an extra processing cost, meaning that language users have limited resources to manage the writing process, as compared to speaking. Moreover, another feature in common with Bourdin and Fayol's study is the amount of revision taking place in written language. These researchers argue that because learners were not allowed to revise written language, the quality of written language may have suffered due to insufficient time to manage written production. In our study, revision was allowed in both tasks. Our twin learners were able to go back to rephrase what they wished to say or to edit what they wrote. However, judging by their written products, the two learners did not seem to make use of the opportunity to revise their writing and they seemed to write as if they were speaking, that is, not adjusting their production to better suit the written mode. Their written texts had many grammatical errors, and the focus seemed to be mainly on using the right words instead of composing complex and accurate sentences.

This leads to another possible explanation for our findings. The language proficiency of these two learners was rather low so that in speaking they had difficulty retrieving the words they sought to use in real time. Therefore, in order to fully express what they wished to explain, they likely circumscribed the idea, resulting in longer sentences, especially with dependent clauses and coordinate phrases. This is in line with the results reported by Chan et al. (2014), who found greater lexical diversity in writing than in speaking. During writing, there may be more time for learners to search for the right lexical items so they can express their ideas more concisely. However, this interpretation is not compatible with the findings by Milton and Hopkins (2006), who found that learners of lower English proficiency levels tend to have better phonological than orthographic vocabulary knowledge. Instead, our finding that low-intermediate learners' lexical use, rather than their syntactic complexity, benefits more from the extra time allowed for revision in writing is in line with the results reported by Verspoor et al. (2012), who found that low-proficiency learners tended to focus more on acquiring the lexicon than on using complex sentence constructions. Caspi (2010) similarly showed that L2 lexicon seems to develop before L2 syntax.

Our second question concerned the developmental patterns of our learners. As expected from a dynamic usage-based perspective, we saw a great deal of variability within each learner; however, because we controlled for as many variables as we could by tracing identical twins in very similar conditions, we expected less variation in general developmental patterns. General patterns have been found before among less similar language users. For example, Caspi (2010) found that three of the four learners showed similar sequences in the development of their lexicon and syntax. Van Dijk et al. (2011) demonstrated how different Spanish learners of English proceeded through similar sequences in their acquisition of English negative verb phrases. However, when we compared the twin learners in this study on measures of syntactic complexity over time using the HMM, there were some similarities but also clear differences between the learners. Both showed more syntactic complexity in speaking early on, and Gloria's syntactic complexity in speaking remained stable during the study, but Grace showed a shift to more complexity in writing rather than speaking in the second and third stages. This finding is also in line with the finding by Chan et al. (2014), where the two girls demonstrated contrasting lexical developmental patterns. A possible explanation for the dissimilar development of written syntactic complexity between the twins may be related to the competition between lexical and syntactic processing. Whereas one twin focused more on the lexicon, the other focused more on syntax. Further research will have to examine this relationship.

Based on our findings, therefore, we conclude that even identical twins with similar personalities and interests who are exposed to similar input within the same environment may demonstrate different developmental paths. Larsen-Freeman (2006) argues that whilst some variation between learners is attributable to external social factors, variation must also come from the internal restructuring that occurs within the language learning system. From a usage-based perspective, one expects not only that input plays an important role in shaping a learner's language, but also that each individual has to find his/her own way of dealing with input as he or she tries to decipher the meaning and use of words or constructions. Identical twins are no exception to this.

Conclusion

Language development is dynamic, with different subsystems of language developing at different rates and interacting continuously, which may result in different relations between these subsystems over time. One of the main contributions of the current study is that is has shown a way of modeling the development of separate measures of syntactic complexity as well as interactions among them to discover moments at which syntactic skills get reorganized and new developmental stages emerge. These stages, when confirmed with the HMM, can be seen to occur when there is an ascertainable overall difference in the ways in which measures interact with each other. The HMM allowed us to objectively observe and chart the changes in several measures at a time; otherwise, would only be able to subjectively observe the overall pattern (i.e., linear growth or decline) of development.

By using the HMM, the current study established a new method of confirming subjective findings within the dynamic usage-based perspective. If we had simply wished to investigate which modality of production was more complex, we would not have needed to use the HMM. However, what the HMM allowed us to accomplish is to examine the way in which the complexity in writing and speaking changes over time, thereby establishing the study as process- rather than product-oriented. The analyses demonstrated that within the timeframe of the study, syntactic complexity in oral language developed sooner than in written language, a pattern also generally found in the L1. This study also found that the time limit advantage of written language over oral language did not seem to result in higher syntactic complexity. The low-intermediate twin learners in fact elaborated their ideas and expressions more in their oral than written production, which resulted in greater syntactic complexity in three dimensions: mean length of T-unit, dependent clause per T-unit, and coordinate phrase per T-unit. However, none of these three syntactic complexity measures revealed depth of syntactic complexity, such as the number of ideas expressed or coherence of speech and writing.

Taken together, our findings showed that two very similar learners, that is, identical twins living in the same household with the same schooling and very similar exposure to English, showed contrasting developmental patterns of syntactic complexity. An important point to note here comes from Stromswold (2006):

> Researchers who study language acquisition often implicitly assume that when one refers to the role of environmental factors on language development, one is primarily referring to postnatal, psychosocial factors such as the quantity or quality of adult linguistic input that children receive. If psychosocial factors have a large impact on language development, this would support theories that argue that language development is largely the result of children's social and language environments (empiricist/emergentist theories). It would also call into

> question nativist/biological theories that argue that language acquisition is largely the result of children's innate, biological endowment (nativist/biological theories). (p. 341)

Therefore, we can conclude that, despite being monozygotic twins, the sisters, in displaying contrasting developmental patterns, provide further support to empiricist/emergentist theories of language development, albeit in a dynamic sense. Even if the input was almost identical for the learners, each learner had to develop her own learning path. This observation makes us wonder to what extent there actually exist "average" learners who develop in similar manners.

Final revised version accepted 4 June 2014

References

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, *26*, 390–395.

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, *11*, 17–34.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*, 5–35.

Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. *International Journal of Psychology*, *29*, 591–620.

Bourdin, B., & Fayol, M. (1996). Mode effects in a sentence production task. *Current Psychology of Cognition*, *15*, 245–264.

Bourdin, B., & Fayol, M. (2002). Even in adults, written production is still more costly than oral production. *International Journal of Psychology*, *37*, 219–227.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Bulté, B. (2013). *The development of complexity in second language acquisition: A dynamic systems approach*. Unpublished doctoral dissertation. University of Brussels, Belgium.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, K. Folkert, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency* (pp. 21–46). Amsterdam: John Benjamins.

Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). *Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment* [Monograph] . *Modern Language Journal*, *94*(suppl).

Caspi, T. (2010). *A dynamic perspective on second language development*. Unpublished doctoral dissertation. University of Groningen, Netherlands.

Caspi, T., & Lowie, W. (2013). The dynamics of L2 vocabulary development: A case study of receptive and productive knowledge. *Revista Brasiliera De Linguística Aplicada*, *13*, 437–462.

Chan, C., & Lee, L. (2013). Model-based unsupervised spoken term detection with spoken queries. *IEEE Transactions on Audio, Speech & Language Processing*, *21*, 1330–1342.

Chan, H. P., Lowie, W., & deBot, K. (2014). A case study of lexical development of writing and speaking in identical twins. In J. Gao, M. Guérin, V. Allegranzi, M. Rivière, L. Sauwala, C. Simon et al. (Eds.), *Actes des 16èmes Rencontres Jeunes Chercheurs en Sciences du Langage: Modèles et modélisation dans les sciences du langage* (pp. 54–65). Paris: HAL-SHS.

Chan, H., Lowie, W., & deBot, K. (in press). Input outside the classroom and vocabulary development: A dynamic perspective. In M. Reif & J. Robinson (Eds.), *Culture and cognition in bilingualism*. Amsterdam: John Benjamins.

Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *Journal of Educational Research*, *69*, 176–183.

de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic system theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*, 7–21.

Dykstra-Pruim, P. (2003). Speaking, writing, and explicit rule knowledge: Toward an understanding of how they interrelate. *Foreign Language Annals*, *36*, 66–76.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford, UK: Oxford University Press.

Fayol, M. (1999). From on-line management problems to strategies in written composition. In M. Torrance & G. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory effects in text production* (pp. 13–23). Amsterdam: Amsterdam University Press.

Garside, R. (1987). The CLAWS word-tagging system. In R. Garside, G. Leech, & G. Sampson (Eds.), *The computational analysis of English: A corpus-based approach* (pp. 30–41). London: Longman.

Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, & A. McEnery (Eds.), *Corpus annotation: Linguistic information from computer text corpora* (pp. 102–121). London: Longman.

Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning*, *26*, 321–351.

Hayiou-Thomas, M. E. (2008). Genetic and environmental influences on early speech, language and literacy development. *Journal of Communication Disorders*, *41*, 397–408.

Homburg, T. J. (1984). Holistic evaluation of ESL compositions: Can it be validated objectively? *TESOL Quarterly*, *18*, 87–107.

Hudelson, S. (1984). Kan yu ret an rayt en ingles: Children become literate in English as a second language. *TESOL Quarterly*, *18*, 221–238.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. Urbana, IL: The National Council of Teachers of English.
Hunt, K. W. (1970). Do sentences in the second language grow like those in the first? *TESOL Quarterly*, *4*, 195–202.
Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, *4*, 51–69.
Kameen, P. (1979). Syntactic skill and ESL writing quality. In C. Yorio, K. Perkins, & J. Schachter (Eds.), *TESOL '79: The learner in focus* (pp. 343–364). Washington, DC: TESOL.
Larsen-Freeman, D. (1976). *Evidence of the need for a second language acquisition index of development*. Washington, DC: ERIC Clearinghouse.
Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, *12*, 439–448.
Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five chinese learners of English. *Applied Linguistics*, *27*, 590–619.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: Bradford Books.
Levy, R., & Galen, A. (2006). Tregex and Tsurgeon: Tools for querying and manipulating tree data structures. *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (pp. 2231–2234). Retrieved February 11, 2015, from http://idiom.ucsd.edu/∼rlevy/papers/levy-andrew-2006.pdf
Lowie, W., Caspi, T., vanGeert, P., & Steenbeek, H. (2011). Modeling development and change. In M. Verspoor, K. deBot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 22–122). Amsterdam: John Benjamins.
Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*, 474–496.
Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*, 85–104.
McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, *8*, 299–325.
McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, *35*, 13–23.
Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *Canadian Modern Language Review*, *63*, 127–147.
Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA. *Applied Linguistics*, *30*, 555–578.

Novotney, S., Schwartz, R., & Ma, J. (2009). Unsupervised acoustic and language model training with small amounts of labelled data. *ICASSP 2009. IEEE International Conference on Acoustics, Speech and Signal Processing*, 4297–4300.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency. *Applied Linguistics*, *24*, 492–518.

Park, A. S., & Glass, J. R. (2008). Unsupervised pattern discovery in speech. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*, 186–197.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavior genetics* (5th ed.). New York: Worth.

Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, *47*, 101–143.

Quinn, E., & Nation, I. S. P. (2007). *Asian and pacific speed readings for ESL learners (English Language Institute Occasional Publication No.24)*. Wellington, NZ: SLALS Victoria University of Wellington.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings IEEE (Institute of Electrical and Electronics Engineers)*, *77*, 257–286.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning. *International Review of Applied Linguistics in Language Teaching*, *45*, 193–213.

Robinson, B., & Mervis, C. (1998). Disentangling early language development: Modeling lexical and grammatical acquisition using and extension of case-study methodology. *Developmental Psychology*, *34*, 363–375.

Ryan, M. S., & Nudd, G. R. (1993). *TheViterbi algorithm (Technical Report No. RR–238)*. Warwick, UK: Department of Computer Science, University of Warwick.

Segal, N. L. (2010). Twins: The finest natural experiment. *Personality and Individual Differences*, *49*, 317–323.

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, *31*, 532–553.

Stromswold, K. (2006). Why aren't identical twins identical? Genetic, prenatal and postnatal factors. *Cognition*, *101*, 333–384.

Tapia, E. (1993). *Cognitive demand as a factor in interlanguage syntax: A study in topics and texts*. Unpublished doctoral dissertation. Indiana University, Bloomington, IN.

vanDijk, M., Verspoor, M., & Lowie, W. (2011). Variability and DST. In M. Verspoor, K. deBot, & W. Lowie (Eds.), *A dynamic approach to second language development* (pp. 55–84). Amsterdam: John Benjamins.

VanGeert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *Modern Language Journal*, *92*, 179–199.

Verspoor, M., deBot, K., & Lowie, W. (Eds.). (2011). *A dynamic approach to second language development: Methods and techniques*. Amsterdam: John Benjamins.

Verspoor, M., Lowie, W., & vanDijk, M. (2008). Variability in second language development from a dynamic systems perspective. *Modern Language Journal*, *92*, 214–231.

Verspoor, M., Schmid, M. S., & Xu, W. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, *21*, 239–263.

Weigle, S. C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Honolulu: University of Hawaii Press.

Wu, J. R. W. (2012). GEPT and English language teaching and testing in Taiwan. *Language Assessment Quarterly*, *9*, 11–25.

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, *31*, 236–259.

Yuan, F., & Ellis, R. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 167–192). Amsterdam: John Benjamins.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and online planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*, 1–27.

Zhang, Y., & Glass, J. R. (2009). Unsupervised spoken keyword spotting via segmental DTW on Gaussian posteriorgrams. *Proceedings of the 2009 IEEE Workshop on Automatic Speech Recognition & Understanding* (pp. 398–403). Retrieved February 10, 2015, from https://groups.csail.mit.edu/sls/publications/2009/ASRU09_Zhang.pdf

Appendix: Sample Transcripts of Speaking and Coding of Writing

Transcription of speaking (Grace)

Original data

I think I think *[restart]* (laughter) *[unnecessary utterance]* that TV influence *[singular verb is not corrected.]* me a lot because I watch TV everyday. That is my hobby. I do not think that newspaper influence [singular verb is not corrected] me a lot because I do not really like to read the newspaper. I I *[restart or repetition]* think that the newsletter *[repair]* newspaper is really dirty so I do not want to touch it . . .

Transcription

I think that TV influence me a lot because I watch TV everyday. That is my hobby. I do not think that newspaper influence me a lot because I do not really like to read the newspaper. I think that the newspaper is really dirty so I do not want to touch it . . .

Coding of writing (Grace)

Original data

Of course [incorrect punctuation] I have pressure. People who live on the Earth have pressures [lexical error is not corrected]. . . . Employees have much more pressure because they [incomplete sentence] always afraid of being fired by their bosses. Bosses have pressures, too. Because [misuse of conjunction] they need to have their company better.

Coded data

Of course, I have pressure. People who live on the Earth have pressures. . . . Employees have much more pressure because they are always afraid of being fired by their bosses. Bosses have pressures, too because they need to have their company better.