Coreference and antecedent representation across languages

Sol Lago[a], Shayne Sloggett[b], Zoe Schlueter[a], Wing Yee Chow[c], Alexander Williams[a,d], Ellen Lau[a] and Colin Phillips[a]

[a] *Department of Linguistics, University of Maryland*
[b] *Department of Linguistics University of Massachusetts Amherst*
[c] *Department of Linguistics, University College London*
[d] *Department of Philosophy, University of Maryland*

Corresponding author: Sol Lago
marlago@uni-potsdam.de
University of Potsdam
Potsdam Research Institute for Multilingualism
Campus Golm, Haus 2
Karl-Liebknecht-Strasse 24-25
14476 Potsdam

Abstract

Previous studies have shown that speakers of languages such as German, Spanish and French reactivate the syntactic gender of the antecedent of a pronoun in order to license gender agreement. As syntactic gender information is assumed to be stored in the lexicon, this has motivated the claim that pronouns in these languages reactivate the lexical entry of their antecedent noun. In contrast, in languages without syntactic gender such as English, lexical retrieval might be unnecessary. Using eye-tracking while reading, we examined whether antecedent retrieval involves rapid semantic and phonological reactivation. We compared German and English. In German, we found early sensitivity to the semantic, but not to the phonological features of the pronoun's antecedent. In English, readers did not immediately show either semantic or phonological effects specific to coreference. We propose that early semantic facilitation arises due to syntactic gender reactivation, and that antecedent retrieval may vary cross-linguistically depending on the type of information relevant to the grammar of each language.

Keywords: coreference; *German*; *English*; *sentence comprehension*; *eye-tracking.*

1. Introduction

Theories of language comprehension aim to describe how speakers create and navigate representations of meaning during sentence processing. With words such as "*told*" and "*month*", comprehension primarily involves retrieving information stored in the mental lexicon. But with anaphoric pronouns, such as "*his*" and "*you*", comprehension does not end with lexical access. Pronouns also need to be associated with a representation of their referent in a mental model of the discourse. This process draws on the lexical properties of the pronoun, on the syntactic class of the antecedent and also on the structure of discourse. In this paper we ask what can be learned about the relations between lexical access and the retrieval of a referent representation by using reading measures targeted at the process of interpreting pronouns.

We investigate whether the relationship between lexical and discourse information is influenced by grammatical properties that vary cross-linguistically. We compare German and English, which differ in a useful way. In contrast with English, German nouns have grammatical gender (masculine, feminine or neuter), which is a syntactic property that differs from conceptual gender (male, female). For example, although boys are male, the diminutive word for "*boy*" in German, "*Buebchen*", is syntactically neuter and permits a neuter pronoun. Also, inanimate nouns lack conceptual gender, but they are specified for syntactic gender (e.g. "*das Haus*", 'the house.NEUTR'; "*die Jacke*", 'the jacket.FEM').

We assume, together with previous research, that the representation of a pronoun's referent in the discourse does not include the syntactic gender of its antecedent noun. Instead, syntactic gender is associated with the pronoun's linguistic antecedent (e.g.

Garnham, 2001; Cacciari, Carreiras & Barbolini-Cionini, 1997; Frazier, Henstra & Flores d' Arcais, 1996; Garnham, Oakhill, Erlich & Carreiras, 1995). If this is the case, then in order to identify a gender-matching antecedent during coreference, German speakers might need to reaccess information outside their discourse model, such as the antecedent's syntactic gender in the lexicon. English speakers, however, might not. In English, which lacks syntactic gender, the features necessary to identify an antecedent (including conceptual gender) can be fully stored in the discourse representation of the pronoun's referent, obviating the need for lexical retrieval (Cloitre & Bever, 1988; Lucas, Tanenhaus & Carlson, 1990).

Given this cross-linguistic contrast between German and English, we focus on two questions: does antecedent retrieval in comprehension differ between languages with and without syntactic gender, such as German and English? And does the existence of syntactic gender in German result in further differences in what other types of antecedent information, phonological or semantic, are initially reactivated at the pronoun? The answer to these questions will shed some light on the structure of memory for prior discourse, and its relationship to other memory structures, such as the lexicon.

Previous evidence of retrieval of semantic and phonological features during coreference

In languages with syntactic gender, such as German, Italian, and Spanish, pronouns must agree in syntactic gender with their antecedent. Previous research shows that comprehenders are sensitive to this requirement in online processing (Cacciari, Carreiras & Barbolini-Cionini, 1997; Carreiras, Garnham & Oakhill, 1993; Frazier, Henstra & Flores d' Arcais, 1996; Garnham, Oakhill, Erlich & Carreiras, 1995; Meyer &

Bock, 1999). If it is assumed that the syntactic gender of nouns is not represented in the discourse but is only stored as part of the lexical entry, then these facts indicate that comprehenders of these languages must retrieve a lexical representation of the pronoun's antecedent during coreference.

However, it is less clear whether the retrieval of the antecedent's syntactic gender is associated with the reactivation of other lexical information such as semantic and phonological features. Here we will adopt a prominent model of the lexicon, which proposes a distinction between the *lemma* of a word, which comprises its semantic and syntactic features, and its *lexeme*, which includes its form, phonological and orthographic (for review, see Levelt, Roelofs & Meyer, 1999). Since this model allows separate access of lemma and lexeme information during processing, it is an open question whether pronouns reactivate only syntactic and semantic features of their antecedent (lemma) or whether the antecedent's lexeme is reaccessed as well (for discussion see van Gompel & Majid, 2004).

With regard to the reactivation of semantic features, it is important to point out that pronouns could re-activate semantic features of their antecedent through two routes: the lexicon or the discourse model. On any account, pronoun interpretation requires reference to the discourse model, which should result in some semantic information about the antecedent being re-activated. However, semantic re-access at the lexical level could be faster or more automatic, and it should give rise to spreading activation effects to semantically related words in the lexical network (Collins & Loftus, 1975; Forster, 1976; Levelt, Roelofs, & Meyer, 1999; Morton, 1979). In contrast, spreading activation effects may or may not occur at the conceptual or discourse level.

Previous evidence for the rapid reactivation of semantic antecedent features comes from cross-modal lexical decision studies in English, where participants performed a lexical decision after hearing a pronoun embedded in a sentence (Leiman, 1982; Nicol, 1988; Shillcock, 1982). These studies found faster responses for words that were semantically related to the antecedent of the pronoun relative to unrelated words. For instance, Shillcock (1982) presented sentences like (1) auditorily. Participants performed a lexical decision to a visual probe at various points in the sentence (the asterisks mark the points where a visual probe appeared on the screen in different trials). At the offset of the pronoun "*he*", a facilitation effect was obtained: lexical decisions were faster for a word that was semantically related to the pronoun's antecedent "*teacher*" (e.g., "*school*") than for an unrelated word matched in length and frequency (e.g., "*street*"). In contrast, when the pronoun was "*it*" decision times were similar, suggesting that the facilitation of "*school*" in the "*he*" condition was not due to residual activation from the word "*teacher*" at the beginning of the sentence, but rather by antecedent reactivation specifically due to coreference between the pronoun "*he*" and its antecedent.

(1) The teacher* did not board the train, for the* simple reason that it / he* was not going to the South Coast of England.

SCHOOL/STREET

Although these results support rapid reactivation of semantic antecedent features in English, there are several concerns about the cross-modal paradigm. First, semantic facilitation in these studies is highly dependent on the choice of control words (e.g. "*street*") and changes in their lexical properties were later found to eliminate the effect (McKoon, Ratcliff and Ward, 1994; for a rejoinder, see Nicol, Fodor, & Swinney, 1994).

A second concern is that the effect might have been due to task-related strategies, instead of automatic reactivation processes. This is because detecting semantic relationships between words improves participants' performance in the lexical decision task, which might encourage them to strategically focus on semantic relationships to perform better (Neely, 1991). As a result, it is unclear whether facilitation effects in cross-modal studies should carry over to more implicit comprehension studies, where participants' main task is to read sentences for meaning.

Meanwhile, reactivation of phonological antecedent features was found in a production study in German (Schmitt, Meyer & Levelt, 1999). Participants were asked to verbally describe two successive pictures of an object (e.g. a flower). When the two pictures showed the same object in different colors, a pronoun was typically used to refer to the repeated object. Schmitt and colleagues found that if participants were interrupted to perform a lexical decision task when they were about to utter the pronoun, they showed an inhibition effect for words phonologically related to the antecedent. That is, words that shared the same onset with the antecedent of the pronoun (e.g., "*Bluse"* 'blouse' phonologically related to "*Blume"* 'flower') showed longer reaction times than unrelated words. This suggested that the form of the antecedent was re-accessed during the planning of the pronoun such that it interfered with the decision to a phonologically related word.

However, we are not aware of evidence in comprehension that the antecedent's form or lexeme is reactivated during coreference. In production, selecting the appropriate form of a pronoun requires speakers to consult the antecedent's gender. In comprehension, however, the antecedent's form is not relevant for establishing

coreference. In addition, to our knowledge there is no evidence that supports reactivation of the antecedent's form in languages without syntactic gender, such as English. Therefore, one of the goals of the current study was to address whether form reactivation occurs in English and German, with a focus on comprehension instead of production. In what follows, we outline two experiments that aimed to examine phonological and semantic antecedent effects using the same task and a controlled comparison across English and German.

Overview of the experiments

We examined the type of information about a pronoun's antecedent that is retrieved from memory in comprehension. Our experiments focused on the retrieval of semantic and phonological features of the antecedent. We asked whether retrieval differed between languages that have syntactic gender, such as German (Experiment 1), and those that do not, such as English (Experiment 2). We tracked comprehenders' eye-movements during reading because it provides a more naturalistic measure of comprehension than the cross-modal paradigm, it has better temporal resolution and it does not require participants to make conscious decisions about the lexical relationships under investigation.

We devised a new strategy to probe for semantic and phonological reactivation. We reasoned that if pronouns quickly reactivate lexical information about their antecedent, then this information should impact the processing of immediately following words. We varied the type of relationship between the antecedent and the word after the pronoun, which we will refer to as the target word. We manipulated whether the

relationship between the antecedent and the target word was semantic or phonological. This strategy is similar to cross-modal studies in that it examines how antecedent reaccess impacts the processing of incoming lexical material. But it differs in that it requires no secondary task, and it keeps the target word constant across the different experimental conditions, as shown in (2-3).

(2) **Semantic conditions**

a. Pronoun, Related
The maintenance men told *the singer* about a problem. They had broken *his* piano and would have to repair that first.

b. Pronoun, Unrelated
The maintenance men told *the deputy* about a problem. They had broken *his* piano and would have to repair that first.

c. Determiner, Related
The maintenance men told *the singer* about a problem. They had broken *the* piano and would have to repair that first.

d. Determiner, Unrelated
The maintenance men told *the deputy* about a problem. They had broken *the* piano and would have to repair that first.

(3) **Phonological conditions**

e. Pronoun, Related
The maintenance men told *the singer* there would be a delay. They said that *his* sink wouldn't be installed until next month.

f. Pronoun, Unrelated
The maintenance men told *the deputy* there would be a delay. They said that *his* sink wouldn't be installed until next month.

g. Determiner, Related
The maintenance men told *the singer* there would be a delay. They said that *the* sink wouldn't be installed until next month.

h. Determiner, Unrelated
The maintenance men told *the deputy* there would be a delay. They said that *the* sink wouldn't be installed until next month.

In the semantic conditions, we adopted a manipulation similar to the previous cross-modal studies on coreference (Leiman, 1982; Nicol, 1988; Shillcock, 1982). In the related conditions, the antecedent and target word were semantically/associatively related ("*singer-piano*"), whereas in the unrelated conditions they were not ("*deputy-piano*"). However, in contrast with previous studies, the target word, "*piano*", was always held constant across conditions, and relatedness was manipulated by varying the antecedent of the pronoun in the first sentence ("*singer*" vs. "*deputy*"). In the phonological conditions we based our manipulation of the phonological overlap between the antecedent and the target word on a previous production study in German (Schmitt, Meyer & Levelt, 1999). In the related conditions, the antecedent and the target word shared the same onset ("*singer-sink*"), whereas in the unrelated conditions there was no phonological or orthographic overlap ("*deputy-sink*").

We hypothesized that if comprehenders immediately reactivate the lexical semantic and/or phonological features of an antecedent upon reading a coreferential pronoun, then these features should impact their processing of the subsequent word. In particular, in previous eye-tracking studies, lexical semantic association between prime-target pairs resulted in shorter reading times to the target word in early and late reading measures, as well as higher skipping rates (Camblin, Gordon, & Swaab, 2007; Carroll & Slowiaczek, 1986; Morris & Folk, 1998). Therefore, if comprehenders immediately reactivate antecedent lexical semantic information upon processing a pronoun, then the target word "*piano*" should elicit shorter reading times when it is semantically related to the antecedent (e.g., "*singer*") than when it is not (e.g., "*deputy*").

Meanwhile, previous studies have shown that orthographic relationships can produce inhibition effects, resulting in longer reading times, higher skipping rates and more regressive eye-movements to a word that is preceded by an orthographically related word (e.g. "*extra*" when preceded by "*extract*" in "More time was allowed to *extract*/*justify* the extra information that was needed"; Patterson, Alcock, & Liversedge, 2011; Patterson, Liversedge, & Davis, 2009). Therefore, if comprehenders immediately reactivate phonological and/or orthographic features of an antecedent upon processing a pronoun, we should observe an inhibition effect in the phonological conditions: the target word "*sink*" should elicit longer reading times when it is phonologically and/or orthographically related to the antecedent (e.g., "*singer*") than when it is not (e.g., "*deputy*").

Finally, it is important to note that if relatedness effects are specifically due to referential processing (and not, for example, to residual activation from first encounter of the antecedent word), then they should be absent (or substantially reduced) in the absence of a coreferential pronoun. As determiners are less likely to immediately reactivate the antecedent noun, we included the determiner conditions to serve as the control comparisons in both the semantic and phonological materials. If semantic and phonological effects are specifically due to antecedent reactivation, then there should be no difference between related and unrelated target words in the determiner conditions.

2. Experiment 1: German

2.1. Methods

2.1.1. Participants

Participants (n = 60, mean age = 25 years, 46 females) were all native speakers of German and were recruited from the University of Potsdam community. All participants provided informed consent and received either course credit or payment for their participation.

2.1.2. Materials and design

The experimental materials consisted of 64 sets of items, distributed in 8 conditions. Each item consisted of a 2-sentence passage. The second sentence contained a masculine or neuter possessive pronoun followed by the target word. The first sentence introduced the antecedent of the pronoun, which was realized as the direct object of a transitive verb. The antecedent was singular and either masculine or neuter in gender. In contrast, the sentential subject was plural and feminine. This ensured that the pronoun in the second sentence unambiguously referred to the direct object of the first sentence. A sample item set is shown in Table 1 and the full item sets are available in the Supplemental Materials.

We varied whether the antecedent of the pronoun was related or unrelated to the target word. This relationship could be semantic (e.g. *Zeichenlehrer-BILD*, 'drawing teacher-PAINTING') or phonological/orthographic (e.g. *Zeichenlehrer-ZEITUNG*, 'drawing teacher-NEWSPAPER'). In the latter case, the antecedent and target word overlapped in at least the first two characters and phonemes of the target's onset ($mean_{orth}$= 2.88, SD_{orth} = 0.95; $mean_{phon}$= 2.51, SD_{phon} = 1.01) according to the WebCelex database (Baayen, Piepenbrock, & Van Rijn, 1993). For the unrelated conditions, the antecedent of the pronoun was replaced with a word that did not share a semantic or

phonological relationship with the target word (e.g. *Administrator-BILD*, 'administrator-PAINTING' and *Administrator-ZEITUNG*, 'administrator-NEWSPAPER' respectively). Semantic relatedness was determined based on the judgments of two native speakers of German.

Related and unrelated antecedents were matched in lemma log frequency ($mean_{rel}$ = 0.88, SD_{rel} = 0.66; $mean_{unrel}$ = 0.86, SD_{unrel} = 0.65) and length ($mean_{rel}$ = 9.08, SD_{rel} = 2.92; $mean_{unrel}$ = 9.30, SD_{unrel} = 2.85) using the German WebCelex database. To isolate relatedness effects specifically due to coreference from relatedness effects due to priming that stemmed from having read the antecedent noun in the previous sentence context, we also manipulated whether the target word was preceded by a pronoun or a determiner. This resulted in a 2 (related/unrelated) × 2 (semantic/phonological) × 2 (pronoun/determiner) design.

The 64 item sets were divided into 8 lists, such that each list contained exactly one version of each item and 8 items in each condition. Thus, each participant never saw more than one version of the same item. The experiment also contained 72 two-sentence filler items of comparable length and complexity. Filler items contained other kinds of referential expressions and anaphors, such as feminine pronouns.

Semantic conditions

a. Pronoun, Related
/Die Nachbarinnen /mochten den /Zeichenlehrer/**,** der im obersten Stockwerk wohnte. /Sie fanden, /dass sein/ Bild/, an dem/ er in /seiner Freizeit gearbeitet hatte und das jetzt im Hausflur hing, sehr gut geworden war./
The neighbors liked the drawing teacher, who lived on the top floor. They thought that his painting, on which he had worked in his spare time and now hung in the hall, had become very good.

b. Pronoun, Unrelated
Die Nachbarinnen mochten den Administrator der im obersten Stockwerk wohnte. Sie fanden, dass sein Bild**,** an dem er in seiner Freizeit gearbeitet hatte und das jetzt im Hausflur hing, sehr gut geworden war.

... administrator… his painting…

c. Determiner, Related
Die Nachbarinnen mochten den Zeichenlehrer**,** der im obersten Stockwerk wohnte. Sie fanden, dass das Bild, an dem er in seiner Freizeit gearbeitet hatte und das jetzt im Hausflur hing, sehr gut geworden war.

... drawing teacher… the painting…

d. Determiner, Unrelated
Die Nachbarinnen mochten den Administrator der im obersten Stockwerk wohnte. Sie fanden, dass das Bild**,** an dem er in seiner Freizeit gearbeitet hatte und das jetzt im Hausflur hing, sehr gut geworden war.

... administrator… the painting…

Phonological conditions

e. Pronoun, Related
/Die Nachbarinnen /mochten den /Zeichenlehrer/, der im obersten Stockwerk wohnte. /Sie gingen sicher, /dass seine/ Zeitung/ nicht /aus seinem/ Briefkasten geklaut wurde./

The neighbors liked the drawing teacher, who lived on the top floor. They made sure that his newspaper was not stolen out of his mailbox.

f. Pronoun, Unrelated
Die Nachbarinnen mochten den Administrator der im obersten Stockwerk wohnte. Sie gingen sicher, dass seine Zeitung nicht aus seinem Briefkasten geklaut wurde.

... administrator… his newspaper…

g. Determiner, Related
Die Nachbarinnen mochten den Zeichenlehrer, der im obersten Stockwerk wohnte. Sie gingen sicher, dass die Zeitung nicht aus seinem Briefkasten geklaut wurde.

... drawing teacher… the newspaper…

h. Determiner, Unrelated
Die Nachbarinnen mochten den Administrator der im obersten Stockwerk wohnte. Sie gingen sicher, dass die Zeitung nicht aus seinem Briefkasten geklaut wurde.

... administrator… the newspaper…

Table 1. Sample set of an experimental item and delimited analysis regions in Experiment 1 (German). The regions of interest are underlined.

2.1.3. Procedure

Participants were tested individually and eye movements were recorded using a desktop-mounted EyeLink 1000 eyetracker (SR Research, Mississauga, Ontario, Canada) interfaced with a Lenovo Thinkpad PC. The sampling rate was 1000 Hz. Stimuli were displayed on a 22-inch EIZO LCD monitor. Sentences were presented in 14 pt. Times New Roman font. Participants were seated 62 cm from the computer screen. At this distance 4.2 characters subtended approximately 1° of visual arc. Viewing was binocular, but only the right eye was recorded. Each sentence was displayed on a single line.

The experiment was implemented using the Experiment Builder software (SR Research, Mississauga, Ontario, Canada). A calibration procedure was performed at the beginning of each testing session, and re-calibration was carried out between trials as needed. Before the experiment began, each participant was instructed to read for comprehension in a normal manner. The participant triggered the onset of each sentence by fixating on a reference point on the left edge of the computer screen where the first word of the sentence was to appear. Each participant read six practice items before the experimental items were shown. All experimental and filler items were followed by a yes/no comprehension question to ensure that participants were attending to the stimuli. Comprehension questions never alluded to the referential dependency between the pronoun and its antecedent. The order of experimental and filler items was pseudo-randomized such that each experimental item was preceded by at least one filler item. The entire experimental session lasted approximately 45 minutes.

2.1.4. Analysis

The initial stages of data analysis were carried out using Data Viewer (SR Research, Mississauga, Ontario, Canada). Examination of the data revealed that no long duration track losses (e.g., missing data for half a line of text or more) occurred at any time during a trial. Fixations were adjusted vertically only in cases where an entire sequence of fixations comprising at least half of the line fell above or below a line of text (i.e. fixations were never adjusted either horizontally or individually). Lastly, fixations shorter than 40 ms or longer than 1000 ms were excluded (0.65%).

The *target region* consisted of the word following the pronoun. We also analyzed the regions immediately before and after the target region: the *target+1* (spillover) and the *pronoun region,* which consisted of the determiner or pronoun together with the preceding complementizer "*dass*" ('that'). Including fixations to the left of a region of interest is a common procedure for analyzing short regions, because short words such as pronouns are often processed during a fixation close to the left of the word when they are skipped (Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994; van Gompel & Majid, 2004). Table 1 shows the division into regions for a sample item.

Since we were interested in whether the recognition of the target word was influenced by antecedent reactivation, we focused our analyses on early eye-tracking measures at the target word. However, we also report *total time* (the sum of all fixations in a region) in order to capture processing differences that occurred after comprehenders' initial processing of the region of interest. For early measures, we report *single fixation duration* (the duration of readers' first fixation in a region when it is the only fixation in the region), *first fixation duration* (the duration of readers' first fixation in a region,

provided that they did not previously fixate on subsequent text) and *first pass reading times* (the sum of all fixations on a critical region before readers leave it for the first time, either to the left or to the right). Also, given that pronouns elicit a large number of regressive eye movements (Ehrlich & Rayner, 1983; van Gompel & Majid, 2004), and that phonological and semantically related words have previously been found to be skipped more often, we report the *probability of regression* and *probability of skipping*. For all dependent variables, trials in which the region under consideration was skipped (i.e., cases where the dependent fixation measure was 0) were excluded from analyses.

Statistical analyses were carried out with R, an open source programming language and environment for statistical computing (R Development Core Team, 2014), using the *lme4* package (Bates, Maechler, Bolker & Walker, 2014). Reading times were logged and then analyzed with linear mixed effects models. Binomial measures and comprehension accuracy were analyzed using mixed effects logistic regression (Jaeger, 2008). P-values were computed with the *lmerTest* package using Satterthwaite's approximation for denominator degrees of freedom (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2014).

We analyzed the semantic and phonological conditions separately, since they contained different target words. The statistical model included fixed effects of determiner type *(pronoun vs. determiner)*, relatedness *(related vs. unrelated)* and their interaction. Both main effects were coded using orthogonal contrasts. For the determiner type factor, the mean of the pronoun conditions was compared with the mean of the determiner conditions. For the relatedness factor, the mean of the related conditions was compared with the mean of the unrelated conditions.

For the random effects structure of the model, we followed current guidelines in psycholinguistics (Barr, Levy, Scheepers, & Tily, 2013) and initially constructed a maximal model that included random intercepts and slopes for all fixed effects and their interaction. Then, to determine whether the inclusion of random slopes was necessary we compared this maximal model with a model with only by-subject and by-item random intercepts. We performed log-likelihood ratio tests (Baayen, Davidson, & Bates, 2008) in the target region and found that the maximal model did not provide a significantly better fit to the data than the intercept-only model in any of the six eye-tracking measures of interest (semantic conditions: $\chi^2_{(18)} = 8.30$, $p = .96$; phonological conditions: $\chi^2_{(18)} = 11.72$, $p = .81$).[1] Therefore, we adopted the intercept-only model, and applied it to the remaining regions of analysis for consistency. We present the model estimates in log milliseconds ($\hat{\beta}$), their standard error, and *t*- and *p*-values in the tables below.

2.2. Results

Average accuracy in the comprehension questions was 93.5%. Table 2 shows means and standard errors for the reading time measures of interest. Table 3 shows the results of the mixed effects model for the logged reading times. Pairwise comparisons and binomial measures are reported in the text.

[1] The values provided are the averaged χ^2 and *p*-values across all measures of interest.

	Measures			
	Single fixation	First fixation	First pass	Total time
Semantic conditions				
Pronoun				
Pronoun, Related	286 (8)	250 (5)	382 (9)	478 (12)
Pronoun, Unrelated	275 (7)	241 (4)	363 (8)	476 (12)
Determiner, Related	263 (5)	242 (4)	320 (7)	391 (9)
Determiner, Unrelated	256 (5)	246 (4)	343 (9)	425 (12)
Target				
Pronoun, Related	222 (6)	226 (5)	311 (12)	353 (13)
Pronoun, Unrelated	242 (6)	239 (5)	329 (11)	399 (14)
Determiner, Related	228 (5)	234 (5)	288 (10)	340 (11)
Determiner, Unrelated	223 (5)	234 (5)	288 (10)	365 (15)
Target+1				
Pronoun, Related	281 (7)	272 (5)	353 (8)	410 (11)
Pronoun, Unrelated	294 (7)	274 (5)	357 (8)	433 (11)
Determiner, Related	283 (6)	267 (5)	345 (9)	404 (11)
Determiner, Unrelated	292 (7)	277 (5)	357 (8)	431 (12)
Phonological conditions				
Pronoun				
Pronoun, Related	278 (8)	244 (4)	371 (9)	463 (12)
Pronoun, Unrelated	271 (7)	247 (4)	382 (9)	469 (11)
Determiner, Related	264 (6)	247 (4)	330 (8)	425 (12)
Determiner, Unrelated	265 (6)	247 (4)	346 (9)	419 (10)
Target				
Pronoun, Related	249 (7)	246 (6)	317 (11)	372 (12)
Pronoun, Unrelated	236 (7)	242 (5)	319 (10)	365 (11)
Determiner, Related	237 (7)	241 (5)	319 (10)	395 (13)
Determiner, Unrelated	239 (6)	242 (5)	322 (11)	378 (15)
Target+1				
Pronoun, Related	281 (6)	270 (5)	351 (9)	424 (12)
Pronoun, Unrelated	264 (4)	259 (4)	342 (9)	396 (11)
Determiner, Related	277 (6)	268 (5)	342 (8)	422 (12)
Determiner, Unrelated	266 (6)	255 (5)	342 (8)	401 (11)

Table 2. Region averages and standard errors in milliseconds in Experiment 1 (German).

	Measures															
	Single fixation				First fixation				First pass				Total time			
	β	SE	*t*	*p*	β	SE	*t*	*p*	β	SE	*t*	*p*	β	SE	*t*	*p*
Semantic conditions																
Pronoun																
Determiner type	**0.07**	**0.02**	**3.68**	**< .01***	-0.01	0.01	-0.43	.67	**0.11**	**0.02**	**5.59**	**< .01***	**0.17**	**0.02**	**8.38**	**< .01***
Relatedness	-0.02	0.02	-1.21	.23	-0.01	0.01	-0.45	.66	0.01	0.02	0.37	.71	0.03	0.02	1.28	.20
Det × Rel	-0.01	0.04	-0.26	.80	-0.04	0.03	-1.49	.14	**-0.10**	**0.04**	**-2.54**	**< .05***	-0.07	0.04	-1.67	.09
Target																
Determiner type	0.01	0.02	0.57	.57	-0.01	0.02	-0.63	.53	**0.07**	**0.02**	**2.87**	**< .01**	0.04	0.02	1.78	.08
Relatedness	**0.04**	**0.02**	**2.02**	**< .05***	0.02	0.02	1.41	.16	0.04	0.02	1.55	.12	**0.08**	**0.02**	**3.21**	**< .01***
Det × Rel	**0.09**	**0.04**	**2.08**	**< .05***	0.05	0.04	1.52	.13	0.07	0.05	1.55	.12	**0.10**	**0.05**	**1.98**	**< .05***
Target+1																
Determiner type	0.00	0.02	-0.20	.84	0.00	0.00	-0.08	.93	0.01	0.02	0.59	.44	0.00	0.02	0.20	0.84
Relatedness	0.03	0.02	1.43	.15	0.00	0.00	1.44	.15	0.02	0.02	1.24	.21	**0.06**	**0.02**	**2.94**	**< .01***
Det × Rel	0.00	0.04	0.13	.89	0.00	0.00	-0.86	.39	-0.03	0.04	-0.91	.36	-0.02	0.04	-0.50	0.62
Phonological conditions																
Pronoun																
Determiner type	0.03	0.02	1.30	.19	-0.01	0.02	-0.70	.49	**0.11**	**0.02**	**5.04**	**< .01***	**0.11**	**0.02**	**5.61**	**< .01***
Relatedness	-0.01	0.02	-0.43	.67	0.00	0.02	-0.11	.91	0.03	0.02	1.56	.12	0.00	0.02	0.09	.93
Det × Rel	-0.02	0.04	-0.54	.59	0.01	0.03	0.20	.84	-0.03	0.04	-0.61	.54	0.00	0.04	-0.08	.94
Target																
Determiner type	-0.01	0.02	-0.47	.64	-0.01	0.02	-0.44	.66	-0.01	0.02	-0.48	.63	-0.02	0.02	-0.95	.34
Relatedness	-0.02	0.02	-0.84	.40	-0.01	0.02	-0.35	.73	-0.01	0.02	-0.29	.77	-0.04	0.02	-1.81	.07
Det × Rel	-0.09	0.05	-1.82	.07	-0.04	0.04	-1.05	.30	0.00	0.05	0.09	.93	0.03	0.05	0.57	.57
Target+1																
Determiner type	0.01	0.02	0.49	0.62	0.01	0.02	0.92	.36	0.01	0.02	0.50	.61	0.01	0.02	0.31	0.76
Relatedness	**-0.06**	**0.02**	**-3.19**	**< .01***	**-0.04**	**0.02**	**-2.78**	**< .01***	-0.02	0.02	-1.18	.24	**-0.06**	**0.02**	**-2.87**	**< .01***
Det × Rel	-0.02	0.04	-0.6	0.56	0.01	0.03	0.21	.84	-0.02	0.04	-0.51	.61	-0.03	0.04	-0.62	0.53

Table 3. Linear mixed-effect model estimates of logged reading times in Experiment 1 (German). For the determiner type factor, a positive estimate indicates that the pronoun conditions were read more slowly than the determiner conditions. For the relatedness factor, a negative estimate indicates that the related conditions were read more quickly than the unrelated conditions. Reliable effects are in bold font.

2.2.1. Semantic conditions

Pronoun region

The pronoun region was skipped on 8.7% of trials. In both early and late measures, there was a main effect of determiner type: pronouns were read more slowly than determiners in single fixation, first pass and total time, and they were also skipped less often ($\hat{\beta}$ = -0.78, *SE* = 0.19, z = -4.18, $p < .01$). These effects are unsurprising since pronouns are less frequent that determiners, and encountering a pronoun should engage additional cognitive processes, such as the search for an antecedent. In addition, there

was an interaction between determiner type and relatedness in first pass reading times; pairwise comparisons revealed that this was due to the determiners in the unrelated conditions being read more slowly than in the related conditions ($\hat{\beta}$ = 0.06, *SE* = 0.03, *t* = 2.14, *p* < .05). There was no difference in the pronoun conditions ($\hat{\beta}$ = -0.05, *SE* = 0.03, *t* = -1.54, *p* = .12).

Target and spillover regions

The target and target+1 regions were skipped on 30.4% and 14.8% of trials respectively. As in the pronoun region, the target region showed a main effect of determiner type in first pass reading times and in probability of regression: target words were read more slowly and elicited fewer regressions when they followed a pronoun than when they followed a determiner ($\hat{\beta}$ = -0.35, *SE* = 0.15, *z* = -2.34, *p* < .05). Additionally, there was a main effect of relatedness: related words elicited fewer regressions ($\hat{\beta}$ = 0.47, *SE* = 0.15, *z* = 3.1, *p* < .01) and shorter single fixation and total reading times than related words. The main effect of relatedness in total reading times persisted in the target+1 region.

Crucially, these main effects were qualified by a significant interaction in both single fixation and total time, which showed that the relatedness effect was driven by the pronoun conditions. Pairwise comparisons revealed that the relatedness effect was only significant in the pronoun conditions: target words were read more quickly when they were semantically related to the antecedent in both single fixation ($\hat{\beta}$ = 0.09, *SE* = 0.03, *t* = 2.74, *p* < .01) and total time ($\hat{\beta}$ = 0.13, *SE* = 0.04, *t* = 3.67, *p* < .01). In contrast, no

difference was observed in the determiner conditions (*single fixation*: $\hat{\beta}$ = -0.00, *SE* = 0.03, *t* = -0.07, *p* = .94; *total time*: $\hat{\beta}$ = 0.02, *SE* = 0.03, *t* = 0.67, *p* = .50).

The same pattern was observed in first fixation and first pass times. Although the interaction term did not reach significance in these measures, we performed pairwise comparisons because they were motivated by the patterns seen in single fixation and total times and by our hypothesis. As expected, semantically related target words were read more quickly than unrelated words in the pronoun conditions in first fixation ($\hat{\beta}$ = 0.05, *SE* = 0.03, *t* = 2.07, *p* < .05) and first pass reading times ($\hat{\beta}$ = 0.07, *SE* = 0.03, *t* = 2.05, *p* < .05). In contrast, no difference was observed in the determiner conditions (*first fixation*: $\hat{\beta}$ = -0.00, *SE* = 0.03, *t* = -0.08, *p* = .93; *first pass*: $\hat{\beta}$ = -0.01, *SE* = 0.03, *t* = -0.18, *p* = .86).

Figure 1 displays semantic facilitation effects as difference scores, which show the difference in mean reading times between the related and unrelated conditions (difference score = $\text{mean}_{\text{related}} - \text{mean}_{\text{unrelated}}$). In early measures, pronouns show clear facilitation with negative difference scores, which reflect shorter reading times for related than unrelated target words. In contrast, there is no sign of facilitation in the determiner condition, where difference scores cluster around 0. In total reading times, both pronouns and determiners show negative difference scores, consistent with facilitation. However, the effect is significantly larger in the pronoun conditions.

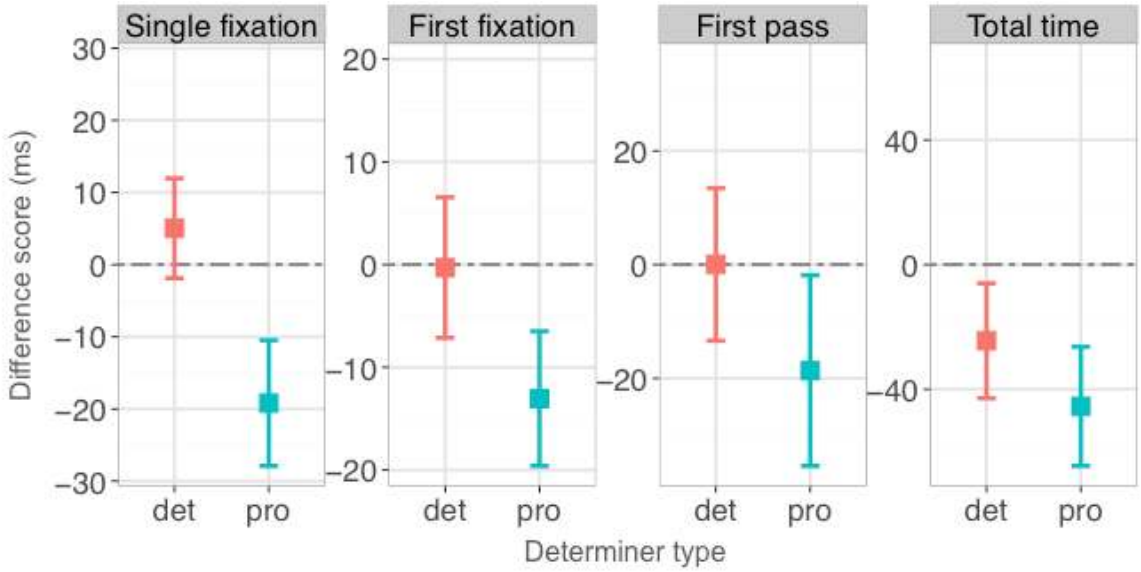

Figure 1. Semantic facilitation effects in the target region in Experiment 1 (German). Mean difference scores and their standard error are shown with squares and bars respectively. Difference scores were computed as the mean difference between the related and unrelated conditions for determiners (*det*) and pronouns (*pro*) separately. Negative difference scores reflect shorter reading times in the related than unrelated conditions. Difference scores are plotted in milliseconds for easier interpretability, but all statistical comparisons were performed on logged reading times. Note that the vertical scales differ because they correspond to different eye-tracking measures.

2.2.2. Phonological conditions

Pronoun region

The pronoun region was skipped on 9.4% of trials. Only a main effect of determiner type was observed in this region: as in the semantic conditions, pronouns were read more slowly than determiners in first pass and total time, and they were also skipped less often ($\hat{\beta}$ = -0.39, *SE* = 0.17, *z* = -2.28, $p < .05$).

Target and spillover regions

The target and target+1 regions were skipped on 26.5% and 14.7% of trials respectively. The only indication of a phonological effect specific to the pronoun

conditions was a marginal interaction between determiner type and relatedness in single fixation duration. However, pairwise comparisons failed to reveal any significant effect of relatedness in either the pronoun ($\hat{\beta}$ = -0.05, *SE* = 0.03, *t* = -1.46, *p* = .15) or the determiner conditions ($\hat{\beta}$ = 0.02, *SE* = 0.03, *t* = 0.73, *p* = .47). Therefore, the pattern of results does not support any effect of the antecedent's form on the reading of the target word.

In the *target+1* region, the phonologically related conditions displayed longer reading times than the unrelated conditions, consistent with an inhibition effect. The main effect of relatedness was significant in single fixation, first fixation and total time and marginal in the probability of regression ($\hat{\beta}$ = -0.26, *SE* = 0.15, *z* = -1.75, *p* < .09). Crucially, there was no interaction between relatedness and pronoun type, suggesting that inhibition affected pronouns and determiners alike.

2.3. Discussion

We examined whether German comprehenders reactivate semantic and phonological antecedent information upon reading a pronoun. We found that the target word after the pronoun was read more quickly when it was semantically related to the pronoun's antecedent than when it was semantically unrelated. In contrast, comprehenders showed no sensitivity to the antecedent phonological features. An inhibition effect consistent with phonological relatedness was only found in the *target+1* region and it occurred for both pronoun and determiners. This suggests that inhibition was due to residual activation from the phonologically related antecedent, and not to reactivation of its form specifically due to the processing of the pronoun.

Crucially, semantic facilitation only occurred when the pre-target word was a pronoun, as supported by a significant relatedness × determiner type interaction. Pairwise comparisons showed that semantically related targets were read more quickly than unrelated targets in the pronoun conditions, but not in the determiner conditions. This pattern suggests that semantic facilitation was specifically due to the processing of coreference.[2] Together with the lack of evidence of a phonological effect specific to the pronoun conditions, these results suggest that German pronouns reactivate semantic but not phonological antecedent information during reading comprehension.

Before we provide an interpretation for the semantic facilitation effect, we should point out that while no facilitation was found for determiners in early measures, the numerical pattern in total reading times was consistent with facilitation: related words after determiners were read on average 25 ms faster than unrelated words. One possibility is that this effect reflects facilitation in later comprehension processes, such as accommodating the meaning of the target word into an ongoing discourse representation. Specifically, readers may have found it easier to incorporate the object "*painting*" to their discourse model when the first sentence introduced a drawing teacher, as opposed to an administrator, as in "*The neighbors liked the drawing teacher/administrator. They thought that the painting, on which he had worked in his spare time had become very good.*" Although the determiner should not have reactivated "*teacher*" initially, the remainder of the sentence supported the interpretation that the painting belonged to the

[2] Although the determiner conditions were intended as a non-referential control, it is possible that at the definite determiner '*the'*, participants expected a continuation that repeated the definite NP referent ("*…the drawing teacher… They made sure that the [drawing teacher]…*"). However, we think this is unlikely because in our materials pronouns were a more felicitous means of referring to the antecedent than were definite descriptions.

drawing teacher/ administrator. To draw this inference might have been easier with "*drawing teacher*" than with "*administrator*", since drawing teachers are more strongly associated to paintings than administrators in the real world. Under this explanation, the facilitation in late reading times for the determiner was due to eased integration processes.

Turning to the semantic facilitation in the pronoun conditions, our results support the hypothesis that when German comprehenders encounter a pronoun, they immediately reactivate the semantic features of its antecedent, which results in the eased recognition of the semantically related target word. One possibility is that the semantic reactivation of the antecedent noun is lexical in nature. As outlined in the Introduction, speakers of languages with syntactic gender might need to retrieve a lexical representation of the antecedent noun during coreference, in order to license antecedent-pronoun gender agreement. If syntactic and semantic features are reaccessed together as part of a word's lemma (e.g. Kempen & Huijbers, 1983; Levelt, 1989) then reaccess of the syntactic gender of the antecedent noun should also reactivate its semantic properties.

Under this account, upon reading the pronoun, German speakers reactivated the lemma of the antecedent noun "*drawing teacher*", which includes its syntactic gender (masculine) and semantic properties. Most models of the lexicon posit that words are stored together in semantic networks, such that activation of a word can spread activation to highly associated words (i.e. a spreading activation mechanism, e.g. Collins & Loftus, 1975; Forster, 1976; Levelt, Roelofs, & Meyer, 1999; Morton, 1979). As a result, target words related to the antecedent noun, such as "*painting*", may have been preactivated

when the antecedent noun was reactivated and they may have been read more quickly later as a result.

A different possibility is that semantic facilitation was due to the reactivation of the pronoun's referent in the discourse model. As comprehenders' mental model of discourse supports their interpretation of a sentence, it should encode some kind of conceptual information. However, it is unclear whether discourse representations can induce spreading activation to semantic associates, as has been proposed for lexical relationships. Although no such semantic-spreading mechanism has been explicitly put forth for discourse models, some authors have suggested that comprehenders can sometimes incorporate concepts that are associated to the pronoun's referent in their discourse model (e.g., Garrod & Terras, 2000).

For example, it is possible that when the antecedent "*drawing teacher*" was encountered in our materials, comprehenders added the concepts of 'student' and 'painting' to their discourse, together with other concepts likely to be present in an event where a drawing teacher is present. When the pronoun was encountered, readers may have reactivated its discourse referent (the drawing teacher introduced in the first sentence) together with related concepts, resulting in the eased recognition of the target word when it matched any of the concepts stored with the pronoun's referent.

Our findings do not unambiguously determine whether the semantic facilitation we observed was due to retrieval of the pronoun's linguistic antecedent or to retrieval of its discourse referent. But since the retrieval of the lexical antecedent was motivated by the presence of syntactic gender in German, we contrasted these possibilities by testing English, a language without syntactic gender. We reasoned that if rapid semantic

facilitation was caused by lexical reactivation due to the existence of syntactic gender, then this effect should be absent in English. In contrast, under a discourse reactivation account, English and German speakers should display similar facilitation effects, as reactivation of the pronoun's referent should occur in both languages.

3. Experiment 2: English

Experiment 2 examined whether English comprehenders show semantic facilitation effects during coreference. A crucial difference between English and German is that English nouns do not have syntactic gender. Instead, gender is either stereotypical ("*janitor*", "*nurse*") or entailed (e.g. "*boy*", "*king*"). In addition, most grammatical noun features such as animacy and number have conceptual correlates: for example, the plural number of a noun usually correlates with the numerosity of its referent in the discourse. Since these features can all arguably be represented in a discourse model, English speakers might not need to retrieve the lexical entry of a pronoun's antecedent because there is no additional benefit or requirement that comes from the antecedent's grammatical information.

We used this cross-linguistic difference to examine the source of the semantic facilitation effect in Experiment 1. We hypothesized that if semantic facilitation was due to reaccess of the syntactic features of the antecedent, then it should not occur in English, where the antecedent's syntactic gender is not grammatically encoded. Alternatively, under an account where spreading activation can occur among related concepts in the discourse (without the need of lexical reactivation) then English comprehenders should

show semantic facilitation effects. An early semantic facilitation effect in English would suggest either that semantic spreading does not require access to the lexicon, or alternatively, that English comprehenders access the lexicon during coreference, despite not needing to.

3.1. Methods

3.1.1. Participants

Participants (n = 60, mean age = 21 years, 38 females) were all native speakers of English and were recruited from the University of Maryland community. All participants provided informed consent and received either course credit or payment for their participation.

3.1.2. Materials and design

We constructed 64 two-sentence item sets in a 2 (related/unrelated) × 2 (semantic/phonological) × 2 (pronoun/determiner) design. The items were based on the antecedent-target pairs in the German experiment, but we prioritized constructing materials that sounded natural in English. Thus, whenever a German noun did not straightforwardly translate to English (e.g. "*Abteilungsleiter*", 'the branch-manager'; "*Rechenkünstler*", 'person who does arithmetic') the antecedent and target nouns were changed and the sentence was adapted.

As in Experiment 1, the possessive pronoun was always in the second sentence and it was singular and had masculine gender. The first sentence introduced the antecedent of the pronoun as the direct object of a transitive verb. In contrast, the subject

in the first sentence was plural and always mismatched the pronoun in number to ensure that all pronouns were unambiguous. The pronoun's antecedent varied in whether it shared a phonological or semantic relationship with the target word. Phonologically related antecedents overlapped with the target word in at least the first two characters and phonemes of the word's onset ($mean_{orth}$ = 2.64, SD_{orth} = 1.03; $mean_{phon}$ = 3.33, SD_{phon} = 0.62). Phonetic transcriptions were obtained from the American pronunciation entries of the Oxford Dictionary and reviewed by a native speaker of American English from the Maryland area.

Semantic relatedness was determined based on the judgments of two native speakers. Related and unrelated antecedents were controlled in log frequency ($mean_{rel}$ = 2.75, SD_{rel} = 0.55; $mean_{unrel}$ = 2.72, SD_{unrel} = 0.61) and length ($mean_{rel}$ = 7.41, SD_{rel} = 1.56; $mean_{unrel}$ = 7.13, SD_{unrel} = 1.60) using the SUBTLex database (Brysbaert & New, 2009; available from the English Lexicon Project, Balota, Yap, Cortese, Hutchison, Kessler, Loftis, Neely, Nelson, Simpson, & Treiman, 2007).

Following Experiment 1, the pronoun region was lengthened by including the complementizer "*that*" before the pronoun. In items without a complementizer, we lengthened the pronoun region by including the last two characters of the preceding verb (26 out of 64 items). The regions of interest for one condition are shown in (4) and a sample item set is shown in (2-3) above. The full item sets are available in the Supplemental Materials.

(4) /The maintenance men/ told the /singer/ about a problem. /They had brok/<u>en his</u>/ <u>piano</u>/ and would/ have/ to repair that first. /

The 64 item sets were divided into 8 lists, such that each list contained exactly one version of each item and 8 items in each condition. Thus, each participant saw each item and each condition, but never saw more than one version of the same item. The experiment also contained 72 two-sentence filler items of comparable length and complexity, which were adapted from the fillers in Experiment 1.

3.1.3. Procedure

Participants were tested individually, and eye movements were recorded using an EyeLink 1000 eyetracker (SR Research, Mississauga, Ontario, Canada), interfaced with a Dell PC. The sampling rate for recordings was 1000 Hz. Stimuli were displayed on a 23-inch Dell LCD monitor. Participants were seated approximately 97 cm from the computer screen. At this distance 6 characters subtended around 1° of visual arc. The angular resolution of the eyetracker was 10–30 min of arc. Viewing was binocular, but only the right eye was recorded. Sentences were presented in 12 pt. fixed width Courier font. Each sentence was displayed on a single line.

The experiment was implemented using the Eye-Track software (http://www.psych.umass.edu/eyelab/software/). A calibration procedure was performed at the beginning of each testing session, and re-calibration was carried out between trials as needed. Each participant was instructed to read for comprehension in a normal manner. The participant triggered the onset of each sentence by fixating on a reference point on the left edge of the computer screen where the first word of the sentence was to appear. Each participant read three practice items before the experimental items were shown. Every experimental and filler item was followed by a yes/no comprehension question to

ensure that participants were attending to the stimuli. Comprehension questions never referred to the referential dependency between the pronoun and its antecedent. The order of experimental and filler items was randomized across participants. The entire experimental session lasted approximately 45 minutes.

3.1.4. Analysis

The initial stages of data analysis were carried out using Eye Doctor (http://www.psych.umass.edu/eyelab/software/). We applied the same processing criteria as in Experiment 1, which resulted in the exclusion of 0.18% of all trials. The same measures of interest and regions of analysis were used.

Following Experiment 1, data from the semantic and phonological conditions were analyzed separately. Statistical analyses were carried out in R with the same statistical model as in Experiment 1, which included fixed effects of determiner type, relatedness and their interaction and random by-subject and by-item intercepts. The intercept-only model was chosen after checking that the maximal model did not provide a significantly better fit to the data in the target region in either of the measures of interest (semantic conditions: $\chi^2_{(18)} = 8.37$, $p = .87$; phonological conditions: $\chi^2_{(18)} = 11.47$, $p = .85$).

3.2. Results

Average accuracy in the comprehension questions was 94.2%. Table 4 shows means and standard errors in the three regions of analysis across the reading time

measures of interest. Table 5 shows the results of the mixed effects model for the logged reading times. Pairwise comparisons and binomial measures are reported in the text.

	Measures			
	Single fixation	First fixation	First pass	Total time
Semantic conditions				
Pronoun				
Pronoun, Related	233 (6)	237 (5)	273 (7)	382 (14)
Pronoun, Unrelated	246 (7)	236 (5)	280 (7)	416 (12)
Determiner, Related	232 (7)	228 (6)	263 (8)	377 (12)
Determiner, Unrelated	235 (7)	225 (4)	257 (7)	383 (11)
Target				
Pronoun, Related	242 (6)	237 (4)	276 (8)	348 (13)
Pronoun, Unrelated	247 (6)	241 (4)	279 (7)	383 (12)
Determiner, Related	239 (5)	241 (4)	275 (7)	363 (11)
Determiner, Unrelated	246 (7)	241 (5)	273 (7)	378 (11)
Target+1				
Pronoun, Related	244 (6)	246 (5)	295 (8)	370 (11)
Pronoun, Unrelated	250 (7)	247 (5)	283 (7)	398 (12)
Determiner, Related	252 (6)	250 (6)	283 (8)	378 (12)
Determiner, Unrelated	245 (6)	243 (5)	279 (6)	396 (12)
Phonological conditions				
Pronoun				
Pronoun, Related	239 (8)	237 (6)	287 (9)	404 (13)
Pronoun, Unrelated	244 (8)	241 (5)	290 (9)	421 (14)
Determiner, Related	223 (7)	223 (5)	257 (8)	399 (14)
Determiner, Unrelated	215 (6)	225 (5)	253 (7)	388 (14)
Target				
Pronoun, Related	251 (6)	244 (4)	286 (7)	418 (14)
Pronoun, Unrelated	248 (7)	244 (5)	295 (8)	413 (14)
Determiner, Related	244 (7)	246 (5)	287 (8)	437 (14)
Determiner, Unrelated	250 (7)	249 (5)	301 (8)	434 (14)
Target+1				

Pronoun, Related	257 (8)	250 (5)	303 (9)	416 (14)
Pronoun, Unrelated	265 (8)	255 (6)	303 (9)	408 (14)
Determiner, Related	247 (7)	242 (5)	297 (8)	436 (15)
Determiner, Unrelated	253 (8)	245 (5)	308 (9)	438 (15)

Table 4. Region averages and standard errors in milliseconds in Experiment 2 (English).

		Measures													
		Single fixation				First fixation				First pass				Total time			
		β	SE	*t*	*p*	β	SE	*t*	*p*	β	SE	*t*	*p*	β	SE	*t*	*p*
Semantic conditions																	
Pronoun																	
	Determiner type	0.02	0.02	1.06	.29	**0.04**	**0.02**	**2.34**	**< .05***	**0.06**	**0.02**	**2.84**	**< .05***	0.04	0.02	1.60	.11
	Relatedness	0.04	0.02	1.72	.09	0.00	0.02	0.00	.99	0.01	0.02	0.26	.80	**0.07**	**0.02**	**2.87**	**< .05***
	Det × Rel	0.07	0.05	1.53	.13	0.02	0.03	0.44	.66	0.06	0.04	1.28	.20	0.08	0.05	1.71	.09
Target																	
	Determiner type	0.01	0.02	0.37	.71	-0.01	0.02	-0.66	.51	0.00	0.02	-0.11	.91	-0.02	0.02	-0.86	.39
	Relatedness	0.02	0.02	0.86	.39	0.01	0.02	0.42	.67	0.00	0.02	0.21	.83	**0.07**	**0.02**	**2.74**	**< .01***
	Det × Rel	0.00	0.04	0.02	.99	0.02	0.03	0.62	.54	0.02	0.04	0.61	.54	0.04	0.05	0.92	.36
Target+1																	
	Determiner type	0.02	0.02	0.69	.49	0.01	0.02	0.61	.54	0.03	0.02	1.51	.13	-0.01	0.02	-0.56	.58
	Relatedness	0.01	0.02	0.36	.72	-0.01	0.02	-0.64	.53	-0.01	0.02	-0.60	.55	**0.06**	**0.02**	**2.55**	**< .05***
	Det × Rel	0.03	0.04	0.74	.46	0.02	0.03	0.44	.66	-0.02	0.04	-0.49	.63	0.02	0.05	0.44	.66
Phonological conditions																	
Pronoun																	
	Determiner type	**0.06**	**0.03**	**2.40**	**< .05***	**0.04**	**0.02**	**2.43**	**< .05***	**0.09**	**0.02**	**3.80**	**< .01***	0.05	0.03	1.76	.08
	Relatedness	0.00	0.03	-0.04	.97	0.02	0.02	1.00	.32	0.01	0.02	0.43	.67	0.01	0.03	0.41	.68
	Det × Rel	0.06	0.05	1.07	.29	0.00	0.04	-0.08	.94	0.02	0.05	0.40	.69	0.06	0.05	1.19	.24
Target																	
	Determiner type	0.02	0.02	0.67	.50	-0.01	0.02	-0.59	.56	-0.01	0.02	-0.65	.51	-0.04	0.03	-1.72	.09
	Relatedness	0.00	0.02	0.18	.86	0.00	0.02	0.31	.76	0.02	0.02	1.24	.22	-0.01	0.03	-0.35	.72
	Det × Rel	-0.04	0.05	-0.81	.42	-0.02	0.03	-0.56	.58	-0.03	0.04	-0.65	.52	-0.03	0.05	-0.60	.55
Target+1																	
	Determiner type	0.04	0.03	1.36	.18	0.03	0.02	1.91	.06	0.01	0.02	0.49	.63	**-0.05**	**0.03**	**-2.08**	**< .05***
	Relatedness	0.02	0.03	0.87	.38	0.02	0.02	1.19	.24	0.03	0.02	1.22	.22	-0.01	0.03	-0.28	.78
	Det × Rel	0.00	0.05	0.04	.97	-0.02	0.04	-0.70	.49	-0.04	0.04	-0.98	.33	-0.02	0.05	-0.32	.75

Table 5. Linear mixed-effect model estimates of logged reading times in Experiment 2 (English). For the determiner type factor, a positive estimate indicates that the pronoun conditions were read more slowly than the determiner conditions. For the relatedness factor, a negative estimate indicates that the related conditions were read more quickly than the unrelated conditions. Reliable effects are in bold font.

3.2.1. Semantic conditions

Pronoun region

The pronoun region was skipped on 28.1% of trials. Early measures showed a main effect of determiner type: pronouns were read more slowly than determiners in first fixation and first pass times, and they were also skipped less often ($\hat{\beta}$ = -0.28, *SE* = 0.11,

z = -2.48, p < .05). Total reading times showed a main effect of relatedness: pronouns and determiners in the related conditions elicited shorter reading times than in the unrelated conditions.

The only interaction that was found between determiner type and relatedness occurred in the probability of regression measure ($\hat{\beta}$ = -0.98, SE = 0.30, z = -3.29, p < .01). Pairwise comparisons showed that the interaction was due to opposite effects of relatedness in the pronoun and determiner conditions: pronouns followed by related target words elicited more regressions than pronouns followed by unrelated words ($\hat{\beta}$ = 0.44, SE = 0.20, z = 2.10, p < .05), and the converse was true for determiners ($\hat{\beta}$ = -0.51, SE = 0.22, z = -2.28, p < .05). This effect was unexpected and we examine it in the Discussion.

Target and spillover regions

The target and target+1 regions were skipped on 21.9% and 25.4% of trials. There were no main effects or interactions in early measures. In total time, related target nouns were read more quickly than unrelated targets across the pronoun and determiner conditions, yielding a main effect of relatedness. Crucially, there was no interaction between relatedness and determiner type. The main effect of relatedness in total reading times persisted in the target+1 region.

To examine the lack of an interaction more closely, we performed a complementary analysis. We computed three eye-tracking measures that are more reflective of late processing: *right bound* (the sum of all first-pass fixation on a region before it is exited to the right), *re-read* (the sum of all first pass fixations on a region after

the region was exited for the first time) and *regression-path times* (the sum of all fixations on a region and the preceding regions before the region of interest is exited to the right). However, no significant interactions between determiner type and relatedness were found in either the target or the spillover region.

Overall, these results show that the semantic relationship between the target word and the antecedent led to faster reading times, but that semantic facilitation occurred for the pronoun and determiner conditions alike. Figure 2 displays the difference scores for comparison with Experiment 1. In early measures, neither pronouns nor determiners show evidence of semantic facilitation and their difference scores cluster around 0. In total reading times, both pronouns and determiners show facilitation, similarly to Experiment 1.

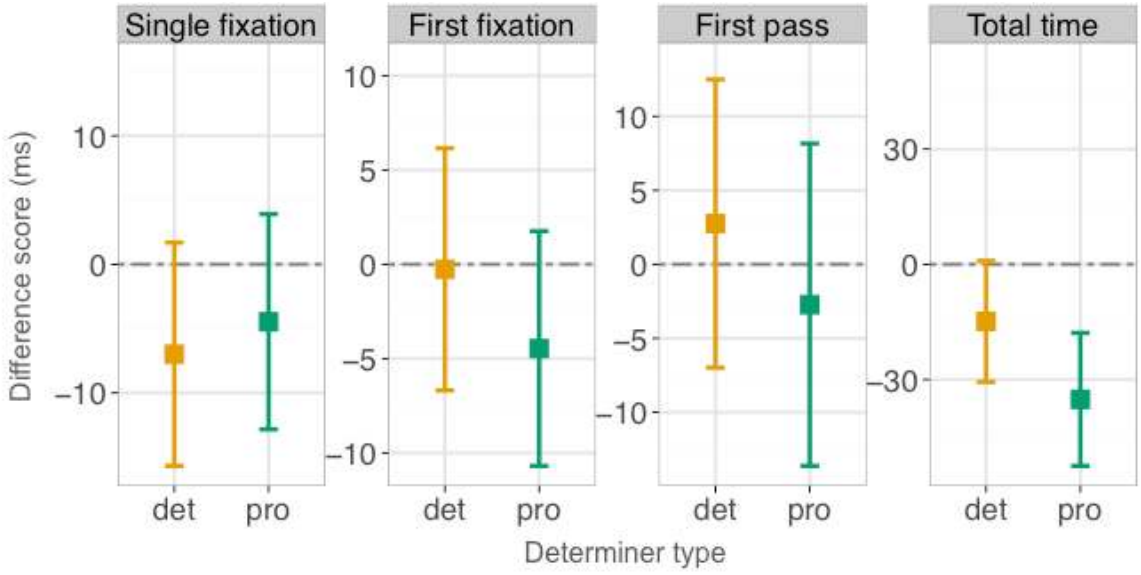

Figure 2. Semantic facilitation effects in the target region in Experiment 2 (English). Mean difference scores and their standard error are shown with squares and bars respectively. Difference scores were computed as the mean difference between the related and unrelated conditions for determiners (*det*) and pronouns (*pro*) separately. Negative difference scores reflect shorter reading times in the related than unrelated conditions. Difference scores are plotted in milliseconds for easier interpretability, but all

statistical comparisons were performed on logged reading times. Note that the vertical scales differ because they correspond to different eye-tracking measures.

3.2.2. Phonological conditions

Pronoun region

The pronoun region was skipped on 30.2% of trials. Pronouns were read more slowly than determiners in single fixation, first fixation and first pass times, and they were also skipped less often ($\hat{\beta}$ = -0.26, *SE* = 0.11, *z* = -2.38, *p* < .05) yielding main effects of determiner type. These effects are likely due to the fact that pronouns were more infrequent that determiners, and that encountering a pronoun should engage additional cognitive processes, such as the search for an antecedent.

Target and spillover regions

The target and target+1 regions were skipped on 19.6% and 27.2% of trials respectively. There were no significant main effects or interactions in the reading time measures. There was a marginal main effect of determiner type in probability of regression: words that followed pronouns elicited more regressions than words that followed determiners ($\hat{\beta}$ = 0.23, *SE* = 0.12, *z* = 1.93, *p* = .05).

The *target+1* region showed a main effect of relatedness: in the related conditions, the spillover region was skipped less often ($\hat{\beta}$ = 0.26, *SE* = 0.12, *z* = 2.21, *p* < .05) and elicited more regressions ($\hat{\beta}$ = -0.28, *SE* = 0.14, *z* = -2.09, *p* < .05). Importantly, there was no interaction between determiner type and relatedness, which suggests similar inhibition effects for pronoun and determiners. Lastly, the target+1 word was read more

quickly in the pronoun than in the determiner conditions, yielding a main effect of determiner type.

3.3. Discussion

We examined whether English speakers show rapid semantic and phonological antecedent reactivation effects during coreference. We reasoned that if the semantic facilitation seen for pronouns in Experiment 1 was due to the existence of syntactic gender, then this effect should be absent in English, which lacks syntactic gender. In contrast, under a discourse reactivation account, English and German speakers should show the same pattern of semantic facilitation, as reactivation of the pronoun's referent occurs in both languages. In addition, we examined whether there was evidence of reactivation of the phonological form of the antecedent during comprehension.

The findings of Experiment 2 contrast with Experiment 1. English comprehenders showed no semantic or phonological antecedent reactivation effects. Unlike German comprehenders, who showed rapid semantic effects in early measures, English comprehenders did not show early differences. However, as in German, later effects of semantic facilitation were observed for both pronouns and determiners in total time. In the phonological conditions, the spillover region showed inhibition in skipping and regression probabilities when it shared an onset with the antecedent noun, but again, these effects impacted both pronouns and determiners.

Taken together, Experiments 1 and 2 show rapid semantic antecedent reactivation during coreference in German, but not in English. This difference supports the hypothesis that facilitation of semantically related words might occur in languages like German

because it is tied to the reaccess of syntactic antecedent features such as grammatical gender. One implication of this view is that in English, reaccess to the referent of the pronoun does not, by itself, reactivate nouns semantically associated to the antecedent noun: for example, the retrieval of the concept of a singer in the discourse does not automatically prime the word "*piano*", as would occur if there were a spreading activation mechanism for discourse. This conclusion will be examined in the General Discussion. In what follows, we discuss two alternative accounts.

One possible explanation for the lack of semantic effects specific to the pronoun conditions in English is that there was a problem in the construction of the antecedent-target noun pairs in English (e.g. "*singer-piano*"). Under this account, the German antecedent-target word pairs may have been more strongly than the English pairs, resulting in the lack of a semantic effect in English. However, this explanation seems unlikely because we did observe overall effects of semantic relatedness in English. In fact, main effects of relatedness were observed across languages in the same measure and with similar magnitude ($\hat{\beta}_{\text{German}} = 0.08$, $SD_{\text{German}} = 0.02$; $\hat{\beta}_{\text{English}} = 0.07$, $SD_{\text{English}} = 0.02$, logged total times). This suggests that the antecedent-target noun pairs successfully elicited meaning associations in English and German. The specific contrast between these languages is that in English semantic effects were not specific to pronouns and occurred only in late reading measures.

A second concern is that the pronoun region was skipped more often in English than in German (28.1% vs. 8.7%). If the reduced number of fixations to pronouns indicates that comprehenders sometimes failed to process them, then antecedent reactivation may not have taken place on some trials thus explaining the absence of

semantic facilitation. However, this explanation seems unlikely for two reasons. The first is that it relies on the assumption that lack of fixations to a region implies lack of processing of that region. But this does not follow, as short words are frequently processed parafoveally (Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994; van Gompel & Majid, 2004). Second, we conducted a supplementary analysis including only the trials where the pronoun region was fixated, and we obtained qualitatively similar patterns: a main effect of relatedness but no interaction between relatedness and determiner type in either the target or the post-target regions. These results suggest that the lack of semantic facilitation in the pronoun conditions was not due to comprehenders' failure to process the pronoun.

Finally, in the semantic conditions we obtained an unexpected interaction between determiner type and relatedness in probability of regression at the pronoun region: there were fewer regressions in the related than in the unrelated pronoun conditions, whereas the converse was true for determiners. This result is surprising, as we did not expect any effect prior to the appearance of the target word. One possibility is that the effect was due to parafoveal preview, if participants' processing of the target word began already at the pronoun region. However, the existence of parafoveal-on-foveal semantic effects is still quite controversial in English (Schotter, Angele, & Rayner, 2012; Rayner, 1998), and it would not explain why the effect was reversed for pronouns and determiners. Therefore, as the effect was not seen in any other measure, and it did not persist to the target or post-target region, we believe that it is more likely to have been

spurious and due to a Type I error.[3] More research will be needed to address this possibility.

4. General Discussion

Our two eye-tracking experiments explored whether pronouns rapidly reactivate lexical semantic and phonological information about their antecedent during comprehension. We examined whether the type of reactivated information depends on the presence of syntactic gender by comparing German, a language with syntactic gender, and English, a language without it. In German, we found early semantic facilitation effects specific to pronouns (Experiment 1) whereas in English we did not (Experiment 2). In contrast, there was no evidence of phonological antecedent reactivation in either of these languages. We discuss each of these profiles in turn.

Semantic effects

Germans comprehenders showed facilitation on early measures when the word after a pronoun was semantically related to its antecedent, while English comprehenders did not. This supports a view where upon encountering a pronoun, German readers reaccess the lemma of the antecedent noun in the lexicon, which includes its syntactic and semantic features. The activation of the antecedent semantic features could in turn preactivate semantically related words, under a spreading activation mechanism (Collins

[3] Since the effect was observed in the probability of regression measure, we examined whether different re-reading behaviors were observed at the antecedent region in the pronoun conditions. However, semantically related antecedents were not more likely to be reread than unrelated antecedents (n.s. main effect of relatedness: $\beta = 0.15$, $SE = 0.14$, $z = 1.07$, $p = .28$) and although they elicited numerically longer reread times than unrelated antecedents, the effect did not reach full significance ($\beta = 0.1$, $SE = 0.05$, $t = 1.95$, $p = .052$).

& Loftus, 1975; Forster, 1976; Levelt, Roelofs, & Meyer, 1999; Morton, 1979). Consistent with this account, we found facilitated processing of the target word across several first-pass reading measures. Crucially, semantic facilitation was not observed in the determiner conditions. Thus, the source of facilitation was likely due to the processing of coreference, and not merely to participants having read semantically related words in the preceding sentence, which was identical in the pronoun and determiner conditions.

In contrast, English speakers showed no evidence of semantic facilitation specific to coreference at the target or spillover words. We propose that this difference is due to the lack of syntactic gender in English. Specifically, coreference in English might not involve lexical retrieval of a pronoun's antecedent, because there is no additional benefit or requirement that comes from the antecedent's grammatical information. In terms of the lexical models outlined previously (e.g. Kempen & Huijbers, 1983; Levelt, 1989; Levelt, Roelofs & Meyer, 1999), English speakers might not reactivate the antecedent's lemma upon encountering a pronoun, which results in the lack of spreading activation to words semantically related to the antecedent noun.

In English, semantic facilitation affected the pronoun and determiner conditions in late reading measures at the target and spillover regions. We suggest that these late effects reflect facilitation in later comprehension processes. Specifically, in sentences such as "*The maintenance men told the singer/deputy about a problem. They had broken his piano and would have to repair that first*", readers may have found it easier to incorporate "*piano*" to their discourse model when the first sentence mentioned a singer instead of a deputy. This is because singers are more related to pianos than deputies in the

real world, such that accommodating the meaning of "*piano*" into an ongoing discourse representation should have been easier in the "singer" case, in both the pronoun and determiner conditions. [4]

Interestingly, our English eye-tracking findings differ from previous cross-modal lexical decision experiments, which did find rapid semantic facilitation effects (Leiman, 1982; Nicol, 1988; Shillcock, 1982). The question is why these studies obtained semantic facilitation to words presented immediately after pronouns, whereas we only observed these effects in late eye-tracking measures. One possibility is that such a contrast is due to a stronger use of explicit strategies in the cross-modal paradigm. In contrast with more implicit paradigms, lexical decision paradigms encourage participants to develop strategic processes, since detecting the semantic relationship between the target word and the antecedent can help them perform better in their lexical decisions (Neely, 1991). Therefore, participants in cross-modal studies may have developed a strategy to focus on semantic antecedent information in order to improve their performance in the task, as opposed to due to automatic reactivation.

Alternatively, the contrast across paradigms might be due to their different temporal resolution. Although both our studies and cross-modal experiments presented target words immediately after pronouns, the time elapsed after the presentation of the pronoun differed between tasks. For instance, in the study by Shillcock (Shillcock, 1982), lexical decisions to words after pronouns took on average 781 ms (*unrelated*: 824 ms;

[4] Note that given the possibility that was raised earlier that the definite determiner predicted coreference to the antecedent NP ("*the singer*…They had broken *the [singer]*…") an alternative account of these data might be that both pronouns and determiners showed semantic facilitation effects due to coreference. Although possible, we find this account unlikely. First, as noted earlier, repeating the earlier definite NP would have been pragmatically strange in many experimental items. Second, in cases where facilitation was observed for both pronouns and determiners, the magnitude of the facilitation effect was always larger for pronouns, which would be unexpected if both pronouns and definite NPs acted to reactivate the antecedent.

related: 738 ms). In our English study, participants spent on average 237 ms reading the pronoun, and 239 ms reading the target word (*first fixation related*: 237 ms; *unrelated*: 241 ms). Thus, our experiment probed for semantic relationships approximately 476 ms after pronoun onset, which constitutes an earlier time window than the cross-modal paradigm. Of course, this estimate is too simplistic because it assumes that latencies in eye-tracking can be obtained by summing first fixation durations. But we provide the estimate, not to establish absolute time points, but instead to suggest that the temporal contrast between tasks might provide access to different stages in pronoun resolution. Ongoing work from our group seems to support this idea, as we have recently found semantic effects at approximately 800 ms after pronoun onset in ERP measures during a sentence comprehension paradigm (Lago et al., in prep).

Finally, we think that the selectivity of our semantic facilitation effects can provide a useful tool to examine the interplay of discourse and lexical information during coreference. Specifically, we found that rapid effects of semantic association are only observed in a language where syntactic gender agreement constraints require comprehenders to retrieve a lexical antecedent representation. In contrast, reaccess of a pronoun's referent in the discourse did not result in rapid semantic association effects in a language without syntactic gender. This suggests that spreading activation of semantic information might only take place in the lexicon. Overall, although both the lexicon and the discourse encode semantic information, the mechanisms used to navigate these levels of representation might be different and might yield differences in information retrieval during comprehension.

Phonological effects

German and English comprehenders showed phonological inhibition effects in the post-target region, with more regressions and longer fixations in the phonologically related conditions. This pattern suggests that our phonological manipulation was able to impact participants' eye movements in both languages. However, phonological inhibition affected the pronoun and determiner conditions alike, which suggests that it was not due to antecedent retrieval. Instead, inhibition may have resulted from residual activation of the antecedent noun. When the antecedent word was read, its orthographic and phonological features should have become activated. If these representations had not fully decayed by the time the target word was read, they may have interfered with the processing of the target word, yielding inhibition effects.

The lack of phonological inhibition in the presence of semantic facilitation in German suggests that not all types of lexical information may be reactivated jointly. For instance, our pattern of results is consistent with the view that pronouns reactivate the lemma but not the lexeme of an antecedent noun during comprehension. This might implicate a difference between the comprehension and the production systems, since in the production of German pronouns, inhibition effects have been previously found for words that were phonologically related to the pronoun's antecedent (Schmitt, Meyer & Levelt, 1999; but see Jescheniak, Schriefers & Hantsch, 2001).

A different explanation for the lack of phonological effects is that the lemma-lexeme distinction corresponds to a difference in the time-course of lexical retrieval. Levelt's model of production proposes a 2-stage sequential architecture, where activation of a word's lemma temporally precedes (and causes) the activation of its lexeme. Thus,

the absence of phonological effects may have arisen if reactivation of the antecedent's form occurred only after readers had fixated the target word. In other words, if inhibition effects are due to co-activation, they might not obtain if the phonology of the antecedent was activated too late, after readers' eyes had already moved to the following word. At present, we cannot address this alternative with eye movement data, but paradigms where the timing of the presentation of the pronoun can be more tightly controlled (e.g. electroencephalography) provide a useful way to address this question.

Conclusion

We used eye movements in reading to examine the retrieval of semantic and phonological antecedent information in German and English. We hypothesized that the existence of syntactic gender in German, but not in English, could influence the type of information retrieved during pronoun comprehension. In German, comprehenders showed evidence of rapid semantic facilitation, in the absence of phonological effects. In contrast, English comprehenders did not show immediate effects of either semantic or phonological antecedent reactivation. We proposed that early semantic facilitation effects might be due to the reactivation of syntactic gender in the lexical entry of a pronoun's antecedent. In contrast, coreference in English might not involve lexical retrieval, because there is no additional benefit or requirement that comes from the antecedent's grammatical information. Taken together, these results suggest that antecedent retrieval during online processing depends on the type of information relevant to the grammar of each language.

Bibliography

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.

Baayen, H. R., Piepenbrock, R., & Van Rijn, H. (1993). The CELEX lexical database. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania. Available from WebCelex at http://celex.mpi.nl/.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. http://CRAN.R-project.org/package=lme4.

Barr D.J., Levy R., Scheepers C., & Tily, H. (2013) Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.

Cacciari, C., Carreiras, M., & Barbolini-Cionini, C. (1997). When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language*, 37, 517–532.

Camblin, C.C., Gordon, P.C., & Swaab, T.Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory & Language*, 56, 103–128.

Carreiras, M., Garnham, A., & Oakhill, J. (1993). The use of superficial and meaning-based representations in interpreting pronouns. Evidence from Spanish. *European Journal of Cognitive Psychology*, 5, 93–116.

Carroll, P., & Slowiaczek, M. L. (1986). Constraints on semantic priming in reading: A fixation time analysis. Memory & Cognition, 14, 509–522.

Cloitre, M. & Bever, T.G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language & Cognitive Processes*, 3, 293–322.

Collins, A.M., & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review,* 82, 407–428.

Ehrlich, K. (1983). Eye movements in pronoun assignment: A study of sentence integration. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes***.** New York: Academic Press.

Forster, K.I. (1976). "Accessing the mental lexicon". In R.J. Wales & E. Walker (eds), *New Approaches to Language Mechanisms*, Amsterdam: North Holland.

Frazier, L., Henstra, J. & Flores d' Arcais, G.B. (1996). Finding candidate antecedents: Phrase or Conceptual Entities. *University of Massachusetts Occasional Papers in Linguistics*, 19,193–238.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. East Sussex: Psychology Press.

Garnham, A., Oakhill, J., Erlich, M. F., & Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: New evidence from Spanish and French. *Journal of Memory and Language,* 34, 41–62.

Garrod, S., Freudenthal, D. & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in discourse. *Journal of Memory and Language*, 33, 39–68.

Jescheniak, J. D., Schriefers, H., & Hantsch, A. (2001). Semantic and phonological activation in noun and pronoun production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1058–1078.

Kempen, G., & Huijbers, P. (1983). The lexicalisation process in sentence production and naming: indirect election of words. *Cognition*, 14, 185–209.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0–6. http://CRAN.R-project.org/package=lmerTest.

Leiman, J.M. (1982). A chronometric analysis of referent assignment to pronouns. Ph.D. dissertation, Wayne State University, Detroit, Michigan.

Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences,* 22*,* 1–75.

Lucas, M.M., Tanenhaus, M.K., & Carlson, G. N. (1990). Levels of representation in the interpretation of anaphoric reference and instrument inference. *Memory and Cognition*, 18, 611–631.

McKoon, G., & Ratcliff, R. (1981). The comprehension processes and memory structures involved in instru- mental inference. *Journal of Verbal Learning and Verbal Behavior,* 20, 671–682.

McKoon, G., Ratcliff, R., & Ward, G. (1994). Testing theories of language processing: An empirical investigation of the on-line lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1219–1228.

Meyer, A. S., & Bock, K. (1999). Representations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory and Language*, 41, 281–30.

Morris, R. K., & Folk, J. R. (1998). Focus as a contextual priming mechanism in reading. *Memory & Cognition*, 26, 1313–1322.

Morton, J. (1979). "Word Recognition". In J. Morton & J.C. Marshall (eds.), *Psycholinguistics Series*, vol. 2, *Structures and Processes.* London: Elek.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (eds.), *Basic processes in reading: Visual word recognition*, (pp. 264–336). Hillsdale, N. J.: Erlbaum.

Nicol, J. (1988). *Coreference Processing during Sentence Comprehension*. Doctoral dissertation, Massachusetts Institute of Technology.

Nicol, J.L., Fodor, J.D. & Swinney, D.A. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1229–1238.

Paterson, K.B., Liversedge, S.P., & Davis, C.J. (2009) Inhibitory neighbor priming effects in eye movements during reading. *Psychonomic Bulletin & Review*, 16, 43–50.

Paterson, K., Alcock, A., & Liversedge, S.P. (2011) Morphological priming during reading: evidence from eye movements. *Language and Cognitive Processes*, 26, 600–623.

Schmitt, B.M., Meyer, A.S. & Levelt, W.J.M. (1999). Lexical access in the production of pronouns. *Cognition*, 69, 313–335.

Schotter, E.R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception & Psychophysics*, 74, 5–35.

Shillcock, R. (1982). The on-line resolution of pronominal anaphora. *Language and Speech*, 25, 385–401.

van Gompel, R.P.G. & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition*, 90, 255–264.

Van Turennout, M., Hagoort, P. & Brown, C. M. (1997) Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 787–806.