Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/17/$12.00 http://dx.doi.org/10.1037/xlm0000343

Coreference and Antecedent Representation Across Languages

Sol Lago
University of Maryland

Shayne Sloggett
University of Massachusetts

Zoe Schlueter
University of Maryland

Wing Yee Chow
University College London

Alexander Williams, Ellen Lau, and Colin Phillips
University of Maryland

Previous studies have shown that speakers of languages such as German, Spanish, and French reactivate the syntactic gender of the antecedent of a pronoun to license gender agreement. As syntactic gender 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. We used eye-tracking while reading to examine 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 varies cross-linguistically depending on the type of information relevant to the grammar of each language.

Keywords: coreference, German, English, sentence comprehension, eye-tracking

Supplemental materials: http://dx.doi.org/10.1037/xlm0000343.supp

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 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 article, we ask what can be learned about the relations between lexical access and the retrieval of discourse referents by using reading measures targeted at the process of interpreting pronouns.

We investigate whether the interaction 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 neutral), 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, *Bübchen*, 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 (Cacciari, Carreiras, & Barbolini-Cionini, 1997; Frazier, Henstra & Flores d'Arcais, 1996; Garnham, 2001; Garnham, Oakhill, Erlich, & Carreiras, 1995). If this is the case, then 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. In contrast, English speakers might not. In English, which lacks syntactic gender, the features necessary to identify an antecedent, which include 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 syntac-

Sol Lago, Department of Linguistics, University of Maryland; Shayne Sloggett, Department of Linguistics, University of Massachusetts; Zoe Schlueter, Department of Linguistics, University of Maryland; Wing Yee Chow, Department of Linguistics, University College London; Alexander Williams, Department of Linguistics and Department of Philosophy, University of Maryland; Ellen Lau and Colin Phillips, Department of Linguistics, University of Maryland.

Correspondence concerning this article should be addressed to Sol Lago, who is now at Potsdam Research Institute for Multilingualism, University of Potsdam, Campus Golm, Haus 2, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany. E-mail: marlago@uni-potsdam.de

tic 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 by a 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 processing (Cacciari, Carreiras, & Barbolini-Cionini, 1997; Carreiras, Garnham, & Oakhill, 1993; Frazier et al., 1996; Garnham, Oakhill, Erlich, & Carreiras, 1995; Meyer & Bock, 1999). For example, Garnham and colleagues (1995) examined whether pronoun resolution is sensitive to the grammatical gender of the antecedent. They used sentences that contained two noun phrases before an object pronoun. Both noun phrases were inanimate to ensure that grammatical gender did not correlate with conceptual gender. In several self-paced reading studies in French and Spanish, reading times for clauses containing pronouns were faster when the two noun phrases differed in grammatical gender, suggesting that speakers were able to use syntactic gender as a cue to disambiguate between the two possible referents. In the domain of language production, Meyer and Bock (1999) found that when participants produced pronouns referring back to antecedents consisting of a noun and a prepositional phrase modifier, they made more agreement errors if the noun inside the prepositional modifier had a different gender than the antecedent noun. This suggests that speakers of languages with syntactic gender retrieve information about the syntactic gender of the antecedents in order to select pronouns.

If it is assumed that the syntactic gender of a noun is not represented in the discourse but only stored under its lexical entry, these previous findings indicate that comprehenders of languages with syntactic gender must retrieve a lexical representation of a pronoun's antecedent during coreference. However, it is less clear whether the retrieval of the antecedent's gender is associated with the reactivation of other lexical information such as semantic and phonological features. Here we will adopt a model of the lexicon that 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). Although this model was originally proposed for production, it has also been fruitfully applied to comprehension partly because it makes specific commitments with regard to how semantic and phonological information are represented and accessed in real time (e.g., Ito, Corley, Pickering, Martin, & Nieuwland, 2016; van Gompel & Majid, 2004). For instance, the model proposes that semantic and phonological information can be accessed separately during lexical retrieval and that phonological access occurs after semantic access. Under the assumption that predicting a word in comprehension is functionally equivalent to generating it in production, Ito and colleagues (2016) showed that the effect of a predictable context (e.g., "The student is going to the library to borrow a . . . ") on the recognition of words related to the expected completion (*book*) was observed rapidly for semantically related words like *page* (within 350 to 450 ms after target word onset) but only after a delay for phonologically related words like *hook* (within 600 to 1,000 ms after target word onset). The authors argued that the delay in phonological over semantic priming supported an account in which comprehenders use their production system to make predictions during comprehension. Meanwhile, van Gompel and Majid (2004) motivated their study on pronoun resolution by proposing that antecedent reactivation in comprehension could be conceptualized as similar to lexical access in production, and they distinguished between accounts in terms of whether pronouns reactivated semantic, syntactic and form-related antecedent features (a *full lexical reaccess account*) or only semantic and syntactic ones (a *lemma re-access account*; van Gompel & Majid, 2004, p. 256).

As the model by Levelt and colleagues (1999) allows separate access to lemma and lexeme information, if the process of antecedent reactivation is similar to lexical access, then pronouns may reactivate either syntactic and semantic features of their antecedent (lemma) or also their form. With regard to the reactivation of semantic features, pronouns could reactivate the 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 reactivated. However, semantic reaccess at the lexical level might 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 et al., 1999; Morton, 1979). In contrast, spreading activation effects may not occur at the 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 task 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 such as the following auditorily:

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

Visual probes: *school/street*

Participants performed a lexical decision to a visual probe at various points in the sentence (the asterisks mark the points in which 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 because of coreference between the pronoun *he* and its antecedent.

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 elimi-

nate the effect (McKoon, Ratcliff, & 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 paradigms, in which participants' main task is to read sentences for comprehension.

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 (1999) 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 response times than unrelated words. This suggested that the form of the antecedent was reaccessed during the planning of the pronoun such that it interfered with the decision to a phonologically related word (but see the research of Jescheniak, Schriefers, & Hantsch, 2001 and Meyer & Bock, 1999, which failed to find an effect of the antecedent's phonology on the production of pronouns using different paradigms).

However, we are not aware of evidence in comprehension that the antecedent's form or lexeme is reactivated during 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 occured 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 antecedent features. 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 to provide a more naturalistic measure of comprehension. In contrast with the cross-modal paradigm, eye-tracking offers a 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 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.

Semantic Conditions

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

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

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

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

Phonological Conditions

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.

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.

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.

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 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 noun 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"; Paterson, Alcock, & Liversedge, 2011; Paterson, 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 of the pronoun (e.g., *singer*) than when it is not (e.g., *deputy*).

To probe whether antecedent reactivation specifically affected the recognition of the target word, our analyses focused on early eye-tracking measures at the target word. We chose single-fixation, first fixation, first pass times, and regression and skipping probabilities, since semantic and phonological effects are often found in these measures. We also report total reading times to provide a measure of second pass processing.

Finally, it is important to note that if relatedness effects are specifically due to referential processing (and not, for instance, 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.

Experiment 1: German

Method

Participants. Participants ($n = 60$; age: $M = 25$ 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.

Materials and design. The experimental materials consisted of 64 sets of items distributed in eight conditions. Each item consisted of a two-sentence passage. The second sentence contained a masculine or neuter possessive pronoun or determiner followed by the target word. Possessive pronouns simultaneously marked the gender of the antecedent and the target word. In German, the stem of a possessive pronoun conveys antecedent agreement (e.g., *Herr Müller . . . sein Sohn* vs. *Frau Müller . . . ihr Sohn*), whereas the suffix of the possessive conveys agreement with the following noun (e.g., *Herr Müller . . . sein Sohn* vs. . . . *seine Tochter*). A sample item set is shown in Table 1, and the full item sets are available in Appendix A of the online supplemental materials.

The first sentence introduced the antecedent of the pronoun, which was realized as the direct object of a transitive verb. Note that in German, grammatical gender is expressed not only on the noun but also on its modifiers (e.g., determiners and adjectives). All antecedents were singular and had male conceptual gender and masculine grammatical gender, with the exception of one antecedent, which was neuter. In contrast, the sentential subjects were always plural and had feminine grammatical gender. This ensured that the pronoun in the second sentence unambiguously referred to the direct object of the first sentence.

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 ($M_{orth} = 2.88$, $SD_{orth} = 0.95$; $M_{phon} = 2.51$, $SD_{phon} = 1.01$) 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 normed in a separate study. Participants ($n = 20$; age: $M = 28$ years; 16 females) were shown the antecedent–target word pairs and asked to rate them on a scale from 1 (*not related*) to 7 (*very related*). Related pairs received a mean rating of 6 ($SD = 1.03$), whereas unrelated pairs received a mean rating of 2.29 ($SD = 1.38$). This difference was statistically significant ($\hat{\beta} = -3.81$, $SE = 0.18$, $t = -20.79$, $p < .05$). In addition, a separate group of participants ($n = 40$; age: $M = 26$ years; 27 females) rated the plausibility of the sentences up to the target word on a scale from 1 (*very implausible*) to 7 (*very plausible*). The conditions in which the antecedent and target words were related were judged as more plausible than the unrelated conditions, but crucially, this difference was similar for the pronoun ($M_{rel} = 5.93$, $SD_{rel} = 1.37$; $M_{unrel} = 4.72$, $SD_{unrel} = 1.76$) and determiner conditions ($M_{rel} = 6.04$, $SD_{rel} = 1.24$; $M_{unrel} = 4.65$, $SD_{unrel} = 1.78$). This resulted in a main effect of semantic relatedness $\hat{\beta} = 1.39$, $SE = 0.18$, $t = 7.93$, $p < .05$) but no interaction between relatedness and determiner type ($\hat{\beta} = -0.18$, $SE = 0.17$, $t = -1.10$, $p = .28$). There were no main effects or interactions in the phonological conditions.

Related and unrelated antecedents were matched in lemma log frequency ($M_{rel} = 0.88$, $SD_{rel} = 0.66$; $M_{unrel} = 0.86$, $SD_{unrel} = 0.65$) and length ($M_{rel} = 9.08$, $SD_{rel} = 2.92$; $M_{unrel} = 9.30$, $SD_{unrel} = 2.85$) 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 one version of each item and eight items in each condition. Thus, each participant never saw more than one version of the same item. The experiment also contained 72 two-sentence

Table 1
Sample Set of an Experimental Item and Analysis Regions in Experiment 1 (German)

Sample item and analysis regions
Semantic conditions
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.
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 . . .
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 . . .
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
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.
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 . . .
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 . . .
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 . . .

Note. The regions of interest were the pre-target, target, and post-target regions, which are delimited between slashes.

filler items of comparable length and complexity. Filler items contained other kinds of anaphors, such as feminine pronouns.

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 1,000 Hz. Stimuli were displayed on a 22-in. 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 recalibration 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 pseudorandomized such that each experimental item was preceded by at least one filler item. The entire experimental session lasted approximately 45 min.

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 in which 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). Last, fixations shorter than 40 ms or longer than 1,000 ms were excluded (0.65% of the data).

The *target region* consisted of the word following the pronoun. We also analyzed the regions immediately before and after the target region: the *post-target* (spillover) and the *pre-target* 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.

Because 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) 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 in which the dependent fixation measure was 0) were excluded from the statistical analyses.

Statistical analyses were carried out with R, an open source programming language and environment for statistical computing (R Development Core Team, 2016), using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). 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 because 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.

In addition to the experimental effects of interest, the analysis included several nuisance variables as covariates to ensure that the results were not due to variability in skipping rates across conditions or differences between word frequency and length across items. For the pre-target region, the nuisance variables included character length and its interaction with the experimental effects of interest. For the target region the nuisance variables included the length of the pre-target region, the length of the target region, the frequency of the target region, whether or not the pre-target region had been skipped. In addition, to ensure that our critical interaction between relatedness and determiner type in single fixation at the target word did not originate from an uneven distribution of fixations across items, the interaction between pre-target skipping rates and the experimental effects of interest was added to the model. All numeric predictors were centered before being entered in the analysis.

We present the model estimates in log ms ($\hat{\beta}$), their standard error, and t, z, and p values in the tables below. The random effects structure of the models included random intercepts and slopes for both experimental effects by participants and items. When the models failed to converge, we gradually simplified them following the guidelines in Barr, Levy, Scheepers, and Tily (2013). In the semantic conditions, nonconvergence occurred for first fixation at the pre-target region and probability of regression at the pre-target and target regions. In the phonological conditions, nonconvergence occurred for first-fixation at the target region and probability of regression at the pre-target and target regions. The simplified models removed the by-item slopes.

Results

Average accuracy in the comprehension questions was 93.4% in the semantic conditions and 93.5% in the phonological conditions. Table 2 shows means and standard errors for the measures of interest. Pairwise comparisons are reported in the text.

Semantic conditions.

Pre-target region. Table 3 shows the mixed effects model results for the pre-target region.[1] The pre-target region was skipped on 8.7% of trials. The only effect at this region was a main effect of determiner type in skipping rates: pronouns were skipped less often than determiners. This effect is likely due to the fact that pronouns were less frequent than determiners, and encountering a pronoun should engage additional cognitive processes, such as the search for an antecedent.

Target and post-target regions. Table 4 shows the mixed effects model results for the target region, and Table 5 shows the results for the post-target region. The target and post-target regions were skipped on 30.4% and 14.8% of trials, respectively. The target region showed a significant interaction between relatedness and determiner type in both single fixation and total time. Pairwise comparisons revealed that the interaction was due to a significant relatedness effect in the pronoun conditions, where target words were read more quickly when they were semantically related to the antecedent (single fixation: $\hat{\beta} = 0.09$, $SE = 0.03$, $t = 2.73$, $p < .05$; total time: $\hat{\beta} = 0.13$, $SE = 0.04$, $t = 3.58$, $p < .05$). In contrast, no difference was observed in the determiner conditions (*single fixation*: $\hat{\beta} = -0.00$, $SE = 0.03$, $t = -0.12$, $p = .91$; total time: $\hat{\beta} = 0.01$, $SE = 0.04$, $t = 0.23$, $p = .82$).

The same pattern was observed in first fixation and first pass times. Although the interaction term was only marginally significant in these measures, we performed pairwise comparisons because they were motivated by our hypothesis and by the patterns

[1] Note that the model outputs for the pre-target region do not include an interaction term between word length and determiner type. This is because the model with the interaction did not converge due to the high multicolinearity between these factors.

Table 2
Region Averages and Standard Errors (in Parentheses) in Ms in Experiment 1 (German)

Region	Measure					
	Single fixation	First fixation	First pass	Total time	Probability of regression (%)	Probability of skipping (%)
Semantic conditions						
Pre-target						
Pronoun, related	286 (8)	250 (5)	382 (9)	478 (12)	11 (1)	6 (1)
Pronoun, unrelated	275 (7)	241 (4)	363 (8)	476 (12)	14 (2)	7 (1)
Determiner, related	263 (5)	242 (4)	320 (7)	391 (9)	11 (1)	10 (1)
Determiner, unrelated	256 (5)	246 (4)	343 (9)	425 (12)	11 (1)	12 (1)
Target						
Pronoun, Related	222 (6)	226 (5)	311 (12)	353 (13)	8 (1)	34 (2)
Pronoun, unrelated	242 (6)	239 (5)	329 (11)	399 (14)	13 (2)	29 (2)
Determiner, related	228 (5)	234 (5)	288 (10)	340 (11)	12 (1)	30 (2)
Determiner, unrelated	223 (5)	234 (5)	288 (10)	365 (15)	14 (2)	29 (2)
Post-target						
Pronoun, related	281 (7)	272 (5)	353 (8)	410 (11)	8 (1)	14 (2)
Pronoun, unrelated	294 (7)	274 (5)	357 (8)	433 (11)	10 (1)	16 (2)
Determiner, related	283 (6)	267 (5)	345 (9)	404 (11)	7 (1)	14 (2)
Determiner, unrelated	292 (7)	277 (5)	357 (8)	431 (12)	7 (1)	16 (2)
Phonological conditions						
Pre-target						
Pronoun, related	278 (8)	244 (4)	371 (9)	463 (12)	11 (1)	8 (1)
Pronoun, unrelated	271 (7)	247 (4)	382 (9)	469 (11)	11 (1)	8 (1)
Determiner, related	264 (6)	247 (4)	330 (8)	425 (12)	12 (1)	13 (2)
Determiner, unrelated	265 (6)	247 (4)	346 (9)	419 (10)	10 (1)	9 (1)
Target						
Pronoun, related	249 (7)	246 (6)	317 (11)	372 (12)	12 (1)	28 (2)
Pronoun, unrelated	236 (7)	242 (5)	319 (10)	365 (11)	13 (2)	26 (2)
Determiner, related	237 (7)	241 (5)	319 (10)	395 (13)	14 (2)	25 (2)
Determiner, unrelated	239 (6)	242 (5)	322 (11)	378 (15)	13 (2)	28 (2)
Post-target						
Pronoun, related	281 (6)	270 (5)	351 (9)	424 (12)	11 (1)	14 (2)
Pronoun, unrelated	264 (4)	259 (4)	342 (9)	396 (11)	10 (1)	14 (2)
Determiner, related	277 (6)	268 (5)	342 (8)	422 (12)	12 (1)	16 (2)
Determiner, unrelated	266 (6)	255 (5)	342 (8)	401 (11)	9 (1)	15 (2)

seen in single fixation and total times. As expected, semantically related target words were read more quickly than unrelated words in the pronoun conditions in first fixation ($\hat{\beta} = 0.05$, $SE = 0.03$, $t = 1.83$, $p = .07$) and first pass reading times ($\hat{\beta} = 0.07$, $SE = 0.03$, $t = 2.10$, $p < .05$). In contrast, no difference was observed in the determiner conditions (first fixation: $\hat{\beta} = -0.00$, $SE = 0.00$, $t = -0.16$, $p = .88$; first pass: $\hat{\beta} = -0.01$, $SE = 0.03$, $t = -0.36$, $p = .72$). However, note that these pairwise comparisons do not provide strong evidence for the absence of a relatedness effect in the determiner conditions in first fixation and first pass because the interaction term in the main model was only marginally significant.

In addition, the model results at the target region showed that some of the nuisance variables influenced reading times. Reading times were often longer when the target word was longer or more infrequent. Further, target words elicited more regressions when the pre-target word was skipped. None of these nuisance variables interacted with the experimental effects of interest.

Figure 1 displays semantic facilitation effects as difference scores, which show the difference in mean reading times between the related and unrelated conditions at the target region for each measure of interest (difference score = $M_{\text{unrelated}} - M_{\text{related}}$). Positive scores index priming, that is, faster reading times in related than unrelated conditions. Negative scores index inhibition, that is, slower reading times in related than unrelated conditions. In early measures, pronouns showed positive scores, consistent with priming. In contrast, there was no sign of facilitation in the determiner conditions, in which difference scores clustered around 0. In total reading times, both pronouns and determiners showed priming. However, the effect was significantly larger in the pronoun conditions.

At the post-target region there was only a main effect of relatedness in total times, which was due to the post-target words being read more quickly in the semantically related conditions.

Phonological conditions.

Pre-target region. Table 6 shows the mixed effects model results for the pre-target region. The pre-target region was skipped on 9.4% of trials. No effects were observed in this region.

Target and post-target regions. Table 7 shows the mixed effects model results for the target region, and Table 8 shows the results for the post-target region. The target and post-target regions were skipped on 26.5% and 14.7% of trials respectively. There were no main effects or interactions for the experimental variables of interest at the target region, except where they interacted with nuisance variables. The analysis of the nuisance variables showed that reading times were often longer when the target word was

Table 3
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 1 (German) in the Semantic Conditions at the Pre-target Region

Pre-target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Length pre-target	.01	.04	.38	.71	.07	.04	1.58	.12
Determiner type	.05	.06	.75	.45	.04	.07	.56	.58
Relatedness	.05	.05	.98	.33	−.05	.05	−.95	.35
Length Pre-target × Relatedness	.04	.05	.77	.44	.01	.06	.21	.84
Determiner Type × Relatedness	−.05	.08	−.57	.57	.05	.09	.52	.60
	First fixation				Probability of regression			
Length pre-target	−.02	.03	−.64	.53	.21	.25	.83	.41
Determiner type	.00	.05	−.06	.95	−.01	.44	−.03	.98
Relatedness	−.02	.04	−.65	.52	−.26	.35	−.74	.46
Length Pre-target × Relatedness	−.01	.04	−.25	.80	−.38	.36	−1.05	.30
Determiner Type × Relatedness	.06	.06	.92	.36	.24	.61	.39	.70
	First pass				Probability of skipping			
Length pre-target	−.01	.04	−.25	.80	.65	.37	1.76	.08
Determiner type	.07	.07	1.06	.29	**−1.99**	**.71**	**−2.79**	**<.05***
Relatedness	−.02	.05	−.47	.64	−.99	.57	−1.75	.08
Length Pre-target × Relatedness	.05	.05	.94	.35	−.94	.56	−1.69	.09
Determiner Type × Relatedness	.03	.09	.34	.73	1.44	.98	1.47	.14

Note. 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 positive estimate indicates that the related conditions were read more quickly than the unrelated conditions. Reliable effects at the alpha = .05 level are indexed with an asterisk and in bold font.

longer or more infrequent. Further, target words elicited more regressions when the pre-target word was skipped. Finally, first pass reading times showed an interaction between pre-target skipping rates and relatedness and a three-way interaction between pre-target skipping rates, relatedness, and determiner type. However, when trials were analyzed separately depending on whether the pre-target word had been skipped, there were no main effects or interactions between determiner type and relatedness.

In the post-target region, words after pronouns were skipped less often than words after determiners. In addition, the phonologically related conditions displayed longer reading times than the unrelated conditions in single fixation, first fixation and total time, consistent with an inhibition effect. Crucially, there was no interaction between relatedness and pronoun type, suggesting that inhibition affected pronouns and determiners alike.

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. In contrast, comprehenders showed no sensitivity to the antecedent phonological features. An inhibition effect consistent with phonological relatedness was only found in the post-target 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 the 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 interaction between relatedness and determiner type in single fixation and total reading times. Pairwise comparisons showed that semantically related targets were read more quickly than were 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. 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.

Note that although the determiner conditions were intended as a nonreferential control, it is possible that at the definite determiner *the*, participants expected a continuation that repeated the definite noun phrase referent (. . . *the drawing teacher* . . . They made sure that the [*drawing teacher*] . . .). To address this possibility, we conducted a supplementary analysis that separated cases in which the target word matched the antecedent in gender (thus being consistent with a repeated antecedent completion) from cases in which it mismatched (thus being inconsistent with a repeated antecedent completion). However, there were no significant differences between these groups, suggesting that the unexpected effect of relatedness in the determiner conditions was not due to participants expecting a repeated noun phrase.

Before we provide an interpretation for the semantic facilitation effect, we should point out that whereas 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. This effect may be due to several reasons. The first factor is that determiners can also introduce anaphoric relationships, as

Table 4
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 1 (German) in the Semantic Conditions at the Target Region

Target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Length pre-target	−.03	.03	−.88	.38	.01	.04	.20	.85
Length target	.01	.00	1.67	.10	**.05**	**.01**	**6.77**	**<.05***
Frequency target	**−.07**	**.02**	**−4.68**	**<.05***	**−.09**	**.03**	**−2.99**	**<.05***
Skipping pre-target	.09	.07	1.37	.17	**.28**	**.07**	**3.98**	**<.05***
Determiner type	.10	.06	1.79	.08	.11	.08	1.48	.14
Relatedness	.01	.03	.18	.86	.01	.04	.23	.82
Skipping Pre-target × Determiner Type	−.05	.12	−.40	.69	−.06	.11	−.52	.60
Skipping Pre-target × Relatedness	−.02	.09	−.22	.82	−.12	.10	−1.18	.24
Determiner Type × Relatedness	**−.09**	**.04**	**−2.15**	**<.05***	**−.13**	**.05**	**−2.52**	**<.05***
Skipping Pre-target × Determiner Type × Relatedness	.15	.18	.86	.39	.12	.17	.70	.48
	First fixation				Probability of regression			
Length pre-target	−.01	.03	−.45	.66	.37	.25	1.50	.13
Length target	.01	.00	1.84	.07	**.13**	**.04**	**3.62**	**<.05***
Frequency target	**−.05**	**.01**	**−3.46**	**<.05***	−.19	.15	−1.28	.20
Skipping pre-target	−.04	.05	−.70	.48	**1.78**	**.37**	**4.83**	**<.05***
Determiner type	.03	.05	.73	.47	−.69	.47	−1.48	.14
Relatedness	.00	.03	−.07	.95	−.44	.28	−1.56	.12
Skipping Pre-target × Determiner Type	.09	.09	1.00	.32	.17	.60	.28	.78
Skipping Pre-target × Relatedness	.01	.07	.12	.91	.34	.55	.63	.53
Determiner Type × Relatedness	−.07	.04	−1.73	.08	−.30	.38	−.80	.42
Skipping Pre-target × Determiner Type × Relatedness	.04	.13	.29	.77	−.72	.95	−.75	.45
	First pass				Probability of skipping			
Length pre-target	−.02	.04	−.45	.65	−.22	.19	−1.19	.24
Length target	**.04**	**.01**	**7.13**	**<.05***	**−.34**	**.04**	**−9.20**	**<.05***
Frequency target	**−.08**	**.02**	**−3.18**	**<.05***	**.33**	**.13**	**2.50**	**<.05***
Skipping pre-target	−.03	.07	−.38	.71	−.64	.41	−1.54	.12
Determiner type	.11	.07	1.60	.11	.34	.33	1.04	.30
Relatedness	.00	.04	−.04	.97	.16	.18	.89	.37
Skipping Pre-target × Determiner Type	.06	.11	.50	.62	.23	.66	.35	.73
Skipping Pre-target × Relatedness	−.01	.09	−.06	.95	−.47	.61	−.77	.44
Determiner Type × Relatedness	−.08	.05	−1.69	.09	.06	.24	.26	.79
Skipping Pre-target × Determiner Type × Relatedness	.07	.16	.44	.66	.94	.95	.99	.32

Note. 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 positive estimate indicates that the related conditions were read more quickly than the unrelated conditions. Reliable effects at the alpha = .05 level are indexed with an asterisk and in bold font.

definite noun phrases may carry the presupposition that their referent was previously introduced in the discourse ("I bought a car. The car was big." See, e.g., Heim, 1988). One interesting implication of our findings is that even if determiners are anaphoric, they do not seem to immediately trigger the reactivation of the antecedent's lemma to the same extent as pronouns.

Another possibility is that numeric pattern in total reading times 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 because drawing teachers are more strongly associated to paintings than administrators in the real world. Although the determiner should not have reactivated *teacher* initially, the remainder of the sentence supported the interpretation that the painting belonged to the drawing teacher/administrator. In fact, several items contained a pronoun later in the sentence that referred to the antecedent and made it clear that the object belonged to him (e.g., "They thought that the *painting*, on which *he* had worked in his spare time . . . "). This explanation would also be consistent with the results of the plausibility norming study, which directly asked participants to rate the likelihood of the sentence contexts and the target word. The results of the study showed that the related conditions were rated as more plausible than the unrelated conditions for pronouns and determiners alike.

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 semantically related target words. One possibility is that the semantic reactivation of the antecedent noun is lexical in nature. As outlined in the introductory paragraphs, speakers of languages with syntactic gender might need to retrieve a lexical representation of the antecedent noun during coreference to license antecedent-pronoun gender agreement. If syntactic and semantic features are reaccessed together as part of a word's lemma (e.g., Kempen &

Table 5
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 1 (German) in the Semantic Conditions at the Post-target Region

Post-target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Determiner type	.00	.02	−.20	.84	.00	.02	.21	.83
Relatedness	.03	.02	1.44	.16	**.06**	**.02**	**2.73**	**<.05***
Determiner × Relatedness	.01	.04	.21	.83	−.02	.04	−.48	.63
	First fixation				Probability of regression			
Determiner type	.00	.02	−.01	.99	.01	.24	.06	.95
Relatedness	.02	.02	1.22	.23	.19	.26	.75	.46
Determiner × Relatedness	−.03	.03	−.88	.38	.59	.43	1.38	.17
	First pass				Probability of skipping			
Determiner type	.01	.02	.45	.65	−.21	.19	−1.12	.26
Relatedness	.02	.02	1.19	.24	.23	.20	1.12	.26
Determiner × Relatedness	−.03	.04	−.86	.39	.04	.28	.15	.88

Note. 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 positive estimate indicates that the related conditions were read more quickly than the unrelated conditions. Reliable effects at the alpha = .05 level are indexed with an asterisk and in bold font.

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 et al., 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. Although most theories assume that the discourse model encodes links to some kind of conceptual information, 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 add concepts that are associ-

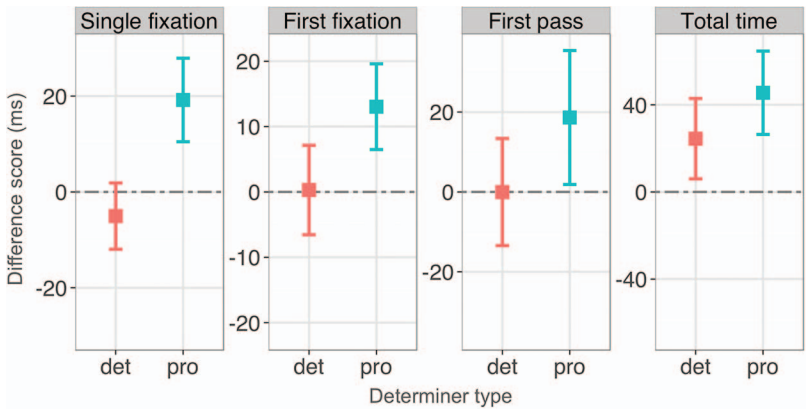

Figure 1. Semantic facilitation effects in the target region in Experiment 1 (German). Mean difference scores and their standard error are shown with squares and bars respectively. Difference scores were computed as the mean difference between the unrelated and related conditions for determiners (*det*) and pronouns (*pro*) separately. Positive scores index priming and negative scores index inhibition. Difference scores are plotted in ms 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. See the online article for the color version of this figure.

Table 6
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 1 (German) in the Phonological Conditions at the Pre-target Region

Pre-target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Length pre-target	.03	.05	.57	.57	.07	.04	1.62	.11
Determiner type	−.02	.07	−.25	.81	.00	.07	−.01	1.00
Relatedness	.06	.05	1.02	.31	−.02	.05	−.38	.71
Length Pre-target × Relatedness	.07	.06	1.11	.27	−.02	.06	−.39	.69
Determiner Type × Relatedness	−.09	.10	−.84	.40	.04	.09	.38	.71
	First fixation				Probability of regression			
Length pre-target	−.02	.03	−.66	.51	.48	.35	1.38	.17
Determiner type	.03	.06	.46	.65	−.55	.59	−.93	.36
Relatedness	.04	.04	1.08	.28	.08	.41	.19	.85
Length Pre-target × Relatedness	.05	.04	1.16	.25	−.24	.43	−.55	.58
Determiner Type × Relatedness	−.08	.07	−1.14	.26	.22	.75	.29	.77
	First pass				Probability of skipping			
Length pre-target	.04	.05	.78	.44	−.22	.41	−.55	.59
Determiner type	.04	.08	.47	.64	.21	.71	.30	.77
Relatedness	−.05	.06	−.94	.35	.19	.52	.36	.72
Length Pre-target × Relatedness	−.01	.06	−.19	.85	−.33	.53	−.63	.53
Determiner Type × Relatedness	.04	.10	.40	.69	−.06	.86	−.07	.95

ated with the pronoun's referent to their discourse model (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 because 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.

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 (e.g., *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 in which 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.

Method

Participants. Participants (n = 60; age: M = 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.

Materials and design. We constructed 64 two-sentence item sets in a 2 (related/unrelated) × 2 (semantic/phonological) × 2 (pronoun/determiner) design. As in Experiment 1, the possessive pronoun always appeared in the second sentence, and it was

Table 7
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 1 (German) in the Phonological Conditions at the Target Region

Target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Length pre-target	−.01	.04	−.22	.83	−.01	.05	−.26	.79
Length target	**.02**	**.01**	**2.89**	**<.05***	**.06**	**.01**	**9.02**	**<.05***
Frequency target	**−.05**	**.02**	**−2.19**	**<.05***	**−.07**	**.03**	**−2.45**	**<.05***
Skipping pre-target	.16	.08	1.91	.06	**.20**	**.08**	**2.38**	**<.05***
Determiner type	−.03	.07	−.35	.73	.02	.09	.19	.85
Relatedness	−.01	.04	−.20	.84	.05	.04	1.40	.16
Skipping Pre-target × Determiner Type	−.13	.12	−1.11	.27	−.09	.12	−.81	.42
Skipping Pre-target × Relatedness	−.14	.11	−1.27	.20	−.03	.10	−.30	.76
Determiner Type × Relatedness	.06	.05	1.30	.19	−.02	.05	−.39	.70
Skipping Pre-target × Determiner Type × Relatedness	.12	.18	.69	.49	.12	.16	.75	.45
	First fixation				Probability of regression			
Length pre-target	.00	.03	−.03	.98	.02	.23	.10	.92
Length target	**.01**	**.00**	**3.37**	**<.05***	.05	.03	1.69	.09
Frequency target	−.03	.02	−1.68	.10	**−.40**	**.14**	**−2.93**	**<.05***
Skipping pre-target	.02	.07	.27	.79	**1.80**	**.45**	**4.03**	**<.05***
Determiner type	−.02	.05	−.30	.77	−.08	.46	−.19	.85
Relatedness	−.01	.03	−.22	.83	.20	.30	.68	.50
Skipping Pre-target × Determiner Type	−.16	.09	−1.78	.08	1.13	.62	1.81	.07
Skipping Pre-target × Relatedness	−.07	.08	−.82	.41	−.36	.57	−.64	.53
Determinater Type × Relatedness	.03	.04	.69	.49	−.08	.36	−.23	.82
Skipping Pre-target × Determiner Type × Relatedness	.17	.13	1.36	.17	−.51	.84	−.61	.54
	First pass				Probability of skipping			
Length pre-target	−.05	.04	−1.29	.20	−.49	.22	−2.25	.02
Length target	**.05**	**.01**	**8.26**	**<.05***	**−.37**	**.04**	**−8.97**	**<.05***
Frequency target	**−.07**	**.03**	**−2.49**	**<.05***	**.33**	**.14**	**2.40**	**<.05***
Skipping pre-target	.01	.09	.17	.86	−.35	.45	−.79	.43
Determiner type	.10	.08	1.23	.22	.58	.39	1.51	.13
Relatedness	.02	.04	.63	.53	−.18	.19	−.92	.36
Skipping Pre-target × Determiner Type	**−.28**	**.12**	**−2.35**	**<.05***	−.30	.68	−.43	.67
Skipping Pre-target × Relatedness	−.13	.11	−1.26	.21	−.23	.61	−.37	.71
Determiner Type × Relatedness	−.03	.05	−.62	.54	.29	.26	1.12	.26
Skipping Pre-target × Determiner Type × Relatedness	**.34**	**.16**	**2.07**	**<.05***	.23	.94	.24	.81

Note. Reliable effects at the alpha = .05 level are indexed with an asterisk and in bold font.

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. We chose antecedent nouns that were likely to be stereotypically masculine to further facilitate referent identification at the pronoun. 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 (M_{orth} = 2.64, SD_{orth} = 1.03; M_{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 normed in a separate study. Participants (n = 20; age: M = 28 years; 10 females) were shown the antecedent–target word pairs and asked to rate them on a scale from 1 (*not related*) to 7 (*very related*). Related pairs received a mean rating of 5.62 (SD = 1.33) whereas unrelated pairs received a mean rating of 2.82 (SD = 1.33). This difference was statistically significant ($\hat{\beta}$ = −3.45, SE = 0.20, t = −17.25, $p < .05$). A separate group of participants (n = 40; age: M = 28 years; 19 females) rated the plausibility of the sentences up to the target word on a scale from 1 (*very implausible*) to 7 (*very plausible*). The conditions in which the antecedent and target words were semantically related were deemed more plausible than the unrelated conditions, but crucially, this difference was similar for the pronoun (M_{rel} = 6.17, SD_{rel} = 1.25; M_{unrel} = 5.23, SD_{unrel} = 1.69) and determiner conditions (M_{rel} = 5.97, SD_{rel} = 1.33; M_{unrel} = 4.16, SD_{unrel} = 1.80). This resulted in a main effect of semantic relatedness ($\hat{\beta}$ = 0.81, SE = 0.19, t = 0.71, $p < .05$) but no interaction between relatedness and determiner type ($\hat{\beta}$ = 0.13, SE = 0.19, t = 0.74, p = .49). There were no main effects or interactions in the phonological conditions.

Related and unrelated antecedents were controlled in log frequency (M_{rel} = 2.75, SD_{rel} = 0.55; M_{unrel} = 2.72, SD_{unrel} = 0.61) and length (M_{rel} = 7.41, SD_{rel} = 1.56; M_{unrel} = 7.13, SD_{unrel} = 1.60) using the SUBTLex database (Brysbaert & New, 2009; available from the English Lexicon Project, Balota et al., 2007).

Table 8
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 1 (German) in the Phonological Conditions at the Post-target Region

Post-target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Determiner type	.01	.02	.42	.67	.01	.02	.26	.80
Relatedness	**−.06**	**.02**	**−2.74**	**<.05***	**−.06**	**.02**	**−2.85**	**<.05***
Determiner × Relatedness	−.02	.04	−.39	.70	−.03	.04	−.61	.54
	First fixation				Probability of regression			
Determiner type	.02	.02	.93	.36	−.06	.19	−.31	.76
Relatedness	**−.05**	**.02**	**−2.65**	**<.05***	−.02	.21	−.08	.94
Determiner × Relatedness	.01	.03	.19	.85	.48	.34	1.38	.17
	First pass				Probability of skipping			
Determiner type	.01	.02	.41	.69	**−.50**	**.23**	**−2.21**	**<.05***
Relatedness	−.02	.02	−1.16	.25	−.09	.21	−.45	.65
Determiner × Relatedness	−.02	.04	−.52	.60	.00	.29	−.01	.99

Note. Reliable effects at the alpha = .05 level are indexed with an asterisk and in bold font.

The gender bias of the antecedent nouns was assessed separately using a gender norming task based on a previous study (Chow, Lewis, & Phillips, 2014). Participants (n = 20; age: M = 28 years; 7 females) were shown the antecedent nouns and asked to rate them on a scale from 1 (*most likely male*) to 7 (*most likely female*). The mean rating of the antecedents was 2.89 (SD = 1.42), indicating that they had a clear male-bias. The mean rating of the antecedent nouns in the related conditions was 2.73 (SD = 1.44) and the mean rating of the antecedent nouns in the unrelated conditions was 3.03 (SD = 1.40). Related and unrelated antecedents did not significantly differ in their gender bias ($\hat{\beta}$ = 0.35, SE = 0.19, t = 1.84, p = .07).

Following Experiment 1, the pre-target region was lengthened by including the complementizer *that* before the pronoun. In items without a complementizer, the pre-target region was lengthened by including the last two characters of the preceding verb (26 out of 64 items). When the post-target region was shorter than 6 characters, it was lengthened by including the immediately following word. The regions of interest for one condition are shown between slashes in the following example:

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

The 64 item sets were divided into eight lists, such that each list contained one version of each item and eight 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. The full item sets are available in the online supplemental materials.

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 1,000 Hz. Stimuli were displayed on a 23-in. Dell LCD monitor. Participants were seated approximately 97 cm from the computer screen. At this distance, six characters subtended around 1° of visual arc. The angular resolution of the eyetracker was 10 to 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 recalibration 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 min.

Analysis. The initial stages of data analysis were carried out using Eye Doctor (http://www.psych.umass.edu/eyelab/software/). We applied the same exclusion criteria and data cleaning procedures 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, the analysis included several nuisance variables as covariates to ensure that the results were not due to variability in skipping rates across conditions or differences between word frequency and length across items. For the pre-target region the nuisance variables included its character length and its interaction with the experimental effects of interest. The use of the region's length as a covariate was necessary because even though pronouns and determiners had equal length (*his* vs. *the*) the pre-target region was lengthened by adding the complementizer *that* (in items that contained it) or the first three character spaces to the left of the pronoun/determiner (in items without a complementizer). Thus, the

length of the pre-target region varied across items and thus it was entered as a factor in the analysis. For the target region the nuisance variables included the length of the pre-target region, the length of the target region, the frequency of the target region, whether the pre-target region had been skipped or not. In addition, to ensure that our critical interaction between relatedness and determiner type in single fixation at the target word did not originate from an uneven distribution of fixations across items, the interaction between pre-target skipping rates and the experimental effects of interest was added to the model. All numeric predictors were centered before being entered in the analysis.

The random effects structure of the models included random intercepts and slopes for both experimental effects by participants and items. When the models failed to converge, we gradually simplified them following the guidelines in Barr, Levy, Scheepers, and Tily (2013). In the semantic conditions, nonconvergence occurred in probability of regression at the target region. The simplified model removed the by-item slopes. In the phonological conditions, nonconvergence occurred in first pass and total reading times at the pre-target region and in single fixation and first fixation in the target region. The simplified models removed the by-item slopes in the former case and the correlation between intercepts and slopes for both items and participants in the remaining cases.

Results

Average accuracy in the comprehension questions was 94.9% in the semantic conditions and 93.3% in the phonological conditions. Table 9 shows means and standard errors in the three regions of analysis across the reading time measures of interest. Pairwise comparisons are reported in the text.

Semantic conditions.

Pre-target region. Table 10 shows the mixed effects model results for the pre-target region. The pre-target 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, first pass, and total reading times, and they also elicited fewer regressions. The probability of regression measure also showed a main effect of relatedness, with more regressions in the related than in the unrelated conditions, and an interaction between determiner type and relatedness. However, none of the pairwise comparisons reached significance (pronouns: $\hat{\beta} = 0.42$, $SE = 0.33$, $z = 1.23$, $p = .19$; determiners: $\hat{\beta} = -0.54$, $SE = 0.31$,

Table 9
Region Averages and Standard Errors (in Parentheses) in Ms in Experiment 2 (English)

Region	Measure					
	Single fixation	First fixation	First pass	Total time	Probability of regression (%)	Probability of skipping (%)
	Semantic conditions					
Pre-target						
Pronoun, related	233 (6)	237 (5)	273 (7)	382 (14)	15 (2)	31 (2)
Pronoun, unrelated	246 (7)	236 (5)	280 (7)	416 (12)	12 (1)	32 (2)
Determiner, related	232 (7)	228 (6)	263 (8)	377 (12)	10 (1)	37 (2)
Determiner, unrelated	235 (7)	225 (4)	257 (7)	383 (11)	13 (2)	37 (2)
Target						
Pronoun, related	242 (6)	237 (4)	276 (8)	348 (13)	17 (2)	23 (2)
Pronoun, unrelated	247 (6)	241 (4)	279 (7)	383 (12)	21 (2)	21 (2)
Determiner, related	239 (5)	241 (4)	275 (7)	363 (11)	18 (2)	22 (2)
Determiner, unrelated	246 (7)	241 (5)	273 (7)	378 (11)	18 (2)	21 (2)
Post-target						
Pronoun, related	244 (6)	246 (5)	295 (8)	370 (11)	7 (1)	27 (2)
Pronoun, unrelated	250 (7)	247 (5)	283 (7)	398 (12)	12 (1)	23 (2)
Determiner, related	252 (6)	250 (6)	283 (8)	378 (12)	10 (1)	27 (2)
Determiner, unrelated	245 (6)	243 (5)	279 (6)	396 (12)	13 (2)	25 (2)
	Phonological conditions					
Pre-target						
Pronoun, related	239 (8)	237 (6)	287 (9)	404 (13)	12 (1)	34 (2)
Pronoun, unrelated	244 (8)	241 (5)	290 (9)	421 (14)	11 (1)	35 (2)
Determiner, related	223 (7)	223 (5)	257 (8)	399 (14)	10 (1)	40 (2)
Determiner, unrelated	215 (6)	225 (5)	253 (7)	388 (14)	11 (1)	41 (2)
Target						
Pronoun, related	251 (6)	244 (4)	286 (7)	418 (14)	23 (2)	22 (2)
Pronoun, unrelated	248 (7)	244 (5)	295 (8)	413 (14)	19 (2)	19 (2)
Determiner, related	244 (7)	246 (5)	287 (8)	437 (14)	17 (2)	20 (2)
Determiner, unrelated	250 (7)	249 (5)	301 (8)	434 (14)	17 (2)	18 (2)
Post-target						
Pronoun, related	257 (8)	250 (5)	303 (9)	416 (14)	17 (2)	26 (2)
Pronoun, unrelated	265 (8)	255 (6)	303 (9)	408 (14)	13 (2)	30 (2)
Determiner, related	247 (7)	242 (5)	297 (8)	436 (15)	20 (2)	24 (2)
Determiner, unrelated	253 (8)	245 (5)	308 (9)	438 (15)	16 (2)	28 (2)

Table 10
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 2 (English) in the Semantic Conditions at the Pre-target Region

Pre-target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Length pre-target	.03	.02	1.32	.19	**.06**	**.02**	**2.44**	**<.05***
Determiner type	.07	.04	1.79	.07	**.08**	**.04**	**2.09**	**<.05***
Relatedness	−.01	.03	−.21	.83	−.03	.04	−.91	.36
Length Pre-target × Determiner Type	**−.06**	**.03**	**−1.99**	**<.05***	−.01	.03	−.21	.84
Length Pre-target × Relatedness	−.03	.03	−1.12	.26	.01	.03	.29	.77
Determiner Type × Relatedness	−.08	.05	−1.72	.09	−.08	.05	−1.68	.09
Length Pre-target × Determiner Type × Relatedness	.04	.04	1.08	.28	−.02	.04	−.55	.58
	First fixation				Probability of regression			
Length pre-target	.03	.02	1.68	.10	**−.29**	**.14**	**−2.07**	**<.05***
Determiner type	**.05**	**.03**	**2.00**	**<.05***	**−.49**	**.23**	**−2.08**	**<.05***
Relatedness	.01	.03	.38	.71	**−.55**	**.24**	**−2.32**	**<.05***
Length Pre-target × Determiner Type	**−.04**	**.02**	**−2.08**	**<.05***	.20	.19	1.03	.30
Length Pre-target × Relatedness	−.03	.02	−1.54	.12	.19	.19	1.01	.31
Determiner Type × Relatedness	−.02	.03	−.57	.57	**.99**	**.30**	**3.30**	**<.05***
Length Pre-target × Determiner Type × Relatedness	.03	.03	.88	.38	−.17	.27	−.65	.52
	First pass				Probability of skipping			
Length pre-target	**.04**	**.02**	**2.07**	**<.05***	**−.49**	**.12**	**−3.96**	**<.05***
Determiner type	**.09**	**.03**	**2.71**	**<.05***	−.21	.18	−1.15	.25
Relatedness	.02	.03	.52	.60	.22	.17	1.30	.19
Length Pre-target × Determiner Type	−.04	.03	−1.66	.10	−.08	.16	−.48	.63
Length Pre-target × Relatedness	.01	.03	.45	.65	.03	.15	.21	.84
Determiner Type × Relatedness	−.05	.04	−1.18	.24	−.24	.23	−1.04	.30
Length Pre-target × Determiner Type × Relatedness	.01	.04	.34	.74	.12	.22	.52	.60

Note. 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 positive estimate indicates that the related conditions were read more quickly than the unrelated conditions. Reliable effects at the alpha = .05 level are indexed with an asterisk and in bold font.

$z = -1.73$, $p = .08$). As the interaction effect was unexpected in the pre-target region, we examine it in more detail in the Discussion section. Finally, the analysis of the nuisance variables showed longer reading times, fewer regressions and lower skipping rates in longer regions. In addition, there was an interaction between determiner type and the length of the pre-target region in single fixation and first fixation times. The interaction was due to the effect of determiner type (i.e., longer reading times for pronouns than determiners) being smaller in longer than shorter regions.

Target and post-target regions. Table 11 shows the mixed effects model results for the target region and Table 12 shows the results for the post-target region. The target and post-target regions were skipped on 21.9% and 25.4% of trials. In early measures there was an interaction between pre-target skipping rates and determiner type at the target region. Pairwise comparisons revealed that the effect of determiner type (i.e., longer reading times for the pronoun than the determiner conditions) was only significant when the pre-target region was skipped. The analysis of the nuisance variables showed that reading times were often longer when the target word was longer or more infrequent. Further, when the pre-target word had been skipped, target words elicited longer total reading times, more regressions and lower skipping rates. Crucially, no interactions between determiner type and relatedness were observed on any measure in the target region.

At the post-target region, related target words elicited longer total reading times and more regressions. Crucially, there was no interaction between relatedness and determiner type, showing that the effect of relatedness affected both pronouns and determiners. 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 priming and their difference scores cluster around 0. In total reading times, both pronouns and determiners show a numeric tendency toward priming, similarly to Experiment 1.

Phonological conditions.

Pre-target region. Table 13 shows the mixed effects model results for the pre-target region. The pre-target region was skipped on 30.2% of trials. Pronouns were read more slowly than determiners in single fixation, first pass and total reading times. The analysis of the nuisance variables revealed that longer regions elicited longer total reading times and lower skipping rates compared with shorter regions.

Target and post-target regions. Table 14 shows the mixed effects model results for the target region and Table 15 shows the results for the post-target region. The target and post-target regions were skipped on 19.6% and 27.2% of trials respectively. There were no experimental effects of interest at the target or post-target regions. The analysis of the nuisance variables at the target region showed that reading times were longer when the target word was

Table 11
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 2 (English) in the Semantic Conditions at the Target Region

Target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Length pre-target	.01	.01	.66	.51	.00	.02	−.01	.99
Length target	.01	.01	.94	.35	**.02**	**.01**	**2.26**	**<.05***
Frequency target	−.03	.02	−1.64	.11	**−.08**	**.03**	**−3.09**	**<.05***
Skipping pre-target	.09	.05	1.72	.09	**.18**	**.05**	**3.28**	**<.05***
Determiner type	.04	.04	1.07	.28	.01	.04	.33	.74
Relatedness	−.02	.04	−.51	.61	−.02	.04	−.52	.60
Skipping Pre-target × Determiner Type	−.06	.07	−.84	.40	−.05	.08	−.61	.54
Skipping Pre-target × Relatedness	.04	.07	.65	.52	−.04	.07	−.48	.63
Determiner Type × Relatedness	.01	.05	.27	.79	−.01	.06	−.21	.83
Skipping Pre-target × Determiner Type × Relatedness	−.08	.10	−.82	.41	−.10	.11	−.94	.35
	First fixation				Probability of regression			
Length pre-target	.01	.01	.87	.39	.01	.07	.09	.93
Length target	.00	.00	.15	.88	−.05	.04	−1.43	.15
Frequency target	−.02	.01	−1.43	.16	−.05	.11	−.45	.65
Skipping pre-target	.07	.04	1.94	.05	**1.39**	**.30**	**4.69**	**<.05***
Determiner type	.04	.03	1.52	.13	.06	.27	.23	.82
Relatedness	.00	.03	.14	.89	.09	.28	.32	.75
Skipping Pre-target × Determiner Type	**−.12**	**.05**	**−2.33**	**<.05***	.09	.41	.22	.83
Skipping Pre-target × Relatedness	.04	.05	.79	.43	−.60	.41	−1.44	.15
Determiner Type × Relatedness	−.03	.04	−.87	.38	−.19	.36	−.53	.60
Skipping Pre-target × Determiner Type × Relatedness	.02	.07	.23	.82	.20	.58	.35	.73
	First pass				Probability of skipping			
Length pre-target	.00	.01	−.18	.86	−.05	.06	−.80	.42
Length target	.01	.01	1.68	.10	**−.26**	**.04**	**−6.44**	**<.05***
Frequency target	**−.05**	**.02**	**−2.73**	**<.05***	**.31**	**.10**	**3.06**	**<.05***
Skipping Pre-Target	.07	.05	1.60	.11	**−1.05**	**.32**	**−3.24**	**<.05***
Determiner type	.04	.04	1.27	.21	.10	.21	.49	.62
Relatedness	.02	.04	.46	.65	.24	.23	1.06	.29
Skipping Pre-target × Determiner Type	−.11	.06	−1.71	.09	−.67	.49	−1.37	.17
Skipping Pre-target × Relatedness	.02	.06	.34	.73	.21	.43	.50	.62
Determiner Type × Relatedness	−.04	.05	−.73	.46	−.04	.29	−.13	.90
Skipping Pre-target × Determiner Type × Relatedness	.02	.09	.20	.84	.26	.64	.41	.68

Note. 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 positive estimate indicates that the related conditions were read more quickly than the unrelated conditions. Reliable effects at the alpha = .05 level are indexed with an asterisk and in bold font.

more infrequent or when the pre-target region was shorter. Further, when the pre-target region had been skipped, target words elicited longer first fixation and total reading times, more regressions and lower skipping rates. Last, the length of the pre-target region and length and frequency of the target regions reliably reduced skipping rates. None of these nuisance variables interacted with the experimental effects of interest.

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 should occur in both languages. In addition, we examined whether there was evidence of reactivation of the phonological form of the antecedent.

The findings of Experiment 2 differ from Experiment 1. English comprehenders showed no early 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, later effects of semantic facilitation were observed for both pronouns and determiners in total reading times and probability of regression at the post-target region. Crucially, the lack of an interaction between semantic facilitation and determiner type suggests that priming 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,

Table 12
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 2 (English) in the Semantic Conditions at the Post-target Region

Post-target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Determiner type	.01	.03	.52	.61	−.02	.03	−.61	.54
Relatedness	.01	.02	.41	.68	**.06**	**.03**	**2.33**	**<.05***
Determiner × Relatedness	.03	.04	.74	.46	.02	.05	.50	.62
	First fixation				Probability of regression			
Determiner type	.01	.02	.57	.57	−.24	.22	−1.10	.27
Relatedness	−.01	.02	−.53	.60	**.50**	**.19**	**2.64**	**<.05***
Determiner × Relatedness	.02	.03	.50	.62	.29	.38	.75	.45
	First pass				Probability of skipping			
Determiner type	.03	.02	1.27	.21	−.15	.14	−1.03	.30
Relatedness	−.01	.02	−.49	.63	−.16	.14	−1.15	.25
Determiner × Relatedness	−.02	.04	−.49	.62	−.09	.24	−.40	.69

Note. 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 positive estimate indicates that the related conditions were read more quickly than the unrelated conditions. Reliable effects at the alpha = .05 level are indexed with an asterisk and in bold font.

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 is examined in the General Discussion. In the following text, 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 associated 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 at the post-target region were observed across languages in the same measure and with similar magnitude. This suggests that the antecedent–target noun pairs successfully elicited meaning associations in English and German. The specific contrast across 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 pre-target 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

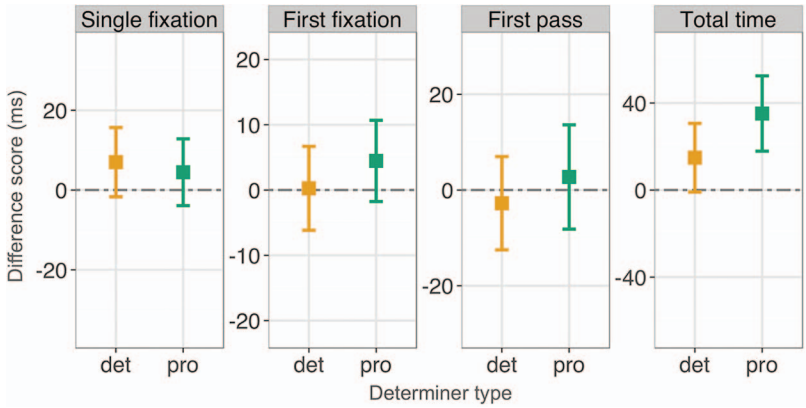

Figure 2. Semantic facilitation effects in the target region in Experiment 2 (English). Mean difference scores and their standard error are shown with squares and bars respectively. Difference scores were computed as the mean difference between the unrelated and related conditions for determiners (*det*) and pronouns (*pro*) separately. Positive scores index priming and negative scores index inhibition. Difference scores are plotted in ms 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. See the online article for the color version of this figure.

Table 13
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 2 (English) in the Phonological Conditions at the Pre-target Region

Pre-target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Length pre-target	.02	.02	.73	.47	**.06**	**.03**	**2.48**	**<.05***
Determiner type	**.09**	**.04**	**2.47**	**<.05***	**.09**	**.04**	**2.19**	**<.05***
Relatedness	.03	.04	.71	.48	.02	.04	.53	.60
Length Pre-target × Determiner type	.00	.03	−.12	.91	−.04	.03	−1.25	.21
Length Pre-target × Relatedness	−.02	.03	−.62	.53	.00	.03	.09	.93
Determiner Type × Relatedness	−.06	.05	−1.09	.28	−.07	.05	−1.24	.22
Length Pre-target × Determiner Type × Relatedness	.03	.04	.71	.48	.04	.04	1.05	.30
	First fixation				Probability of regression			
Length pre-target	.00	.01	.05	.96	−.11	.12	−.86	.39
Determiner type	.04	.03	1.66	.10	−.11	.22	−.51	.61
Relatedness	−.02	.03	−.61	.54	−.39	.25	−1.57	.12
Length Pre-target × Determiner Type	−.01	.02	−.39	.70	.00	.17	.03	.98
Length Pre-target × Relatedness	.00	.02	−.04	.97	−.16	.17	−.92	.36
Determiner Type × Relatedness	.00	.04	.00	1.00	.26	.30	.86	.39
Length Pre-target × Determiner Type × Relatedness	.01	.03	.53	.60	.22	.24	.91	.36
	First pass				Probability of skipping			
Length pre-target	.02	.02	1.00	.32	**−.45**	**.09**	**−5.02**	**<.05***
Determiner type	**.10**	**.03**	**2.90**	**<.05***	−.30	.16	−1.81	.07
Relatedness	.00	.03	.02	.98	.00	.16	.01	.99
Length Pre-target × Determiner Type	.01	.03	.28	.78	.03	.13	.24	.81
Length Pre-target × Relatedness	.00	.03	.14	.89	.08	.12	.65	.52
Determiner Type × Relatedness	−.02	.05	−.42	.67	−.04	.22	−.17	.87
Length Pre-target × Determiner Type × Relatedness	.01	.04	.20	.84	−.08	.18	−.44	.66

Note. Reliable effects at the alpha = .05 level are indexed with an asterisk and in bold font.

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 pre-target region was fixated, and we obtained qualitatively similar patterns with 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, the semantic conditions yielded an unexpected interaction between determiner type and relatedness in probability of regression at the pre-target region: there were numerically fewer regressions in the related than in the unrelated determiner conditions, whereas the converse was true for pronouns. At present, we do not have an explanation for these findings, as the pre-target region did not contain any semantic information, and it is unlikely that participants processed the meaning of the entire target word parafoveally: the existence of parafoveal-on-foveal semantic effects is controversial and there is no evidence that they extend to the semantic meaning of the entire previewed word (Schotter, Angele, & Rayner, 2012; Rayner, 1998). Further, none of the pairwise comparisons in the pronoun and determiner conditions were significant and the effects did not persist in any critical region. Thus, we are unclear about their interpretation and reliability and we believe that more research will be necessary to address this issue.

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 depended 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 in early measures when the word after a pronoun was semantically related to its antecedent. 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 semantic features of the antecedent could in turn preactivate semantically related words,

Table 14
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 2 (English) in the Phonological Conditions at the Target Region

Target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Length pre-target	**−.03**	**.01**	**−2.80**	**<.05***	−.01	.02	−.68	.50
Length target	.00	.01	−.26	.80	.01	.01	1.00	.32
Frequency target	−.02	.02	−.96	.34	**−.12**	**.03**	**−4.40**	**<.05***
Skipping pre-target	.08	.05	1.66	.10	**.15**	**.06**	**2.63**	**<.05***
Determiner type	.00	.05	.09	.92	−.02	.05	−.37	.71
Relatedness	−.03	.04	−.70	.48	.01	.05	.31	.76
Skipping Pre-target × Determiner Type	−.06	.07	−.82	.41	−.12	.08	−1.53	.13
Skipping Pre-target × Relatedness	.01	.07	.09	.93	.00	.08	−.04	.97
Determiner Type × Relatedness	.05	.06	.84	.40	−.02	.06	−.34	.73
Skipping Pre-target × Determiner Type × Relatedness	−.01	.10	−.07	.94	.14	.11	1.25	.21
	First fixation				Probability of regression			
Length pre-target	−.01	.01	−1.64	.10	**.15**	**.07**	**2.22**	**<.05***
Length target	−.01	.00	−1.24	.22	.02	.05	.51	.61
Frequency target	−.02	.01	−1.45	.15	.21	.12	1.80	.07
Skipping pre-target	**.08**	**.04**	**2.16**	**<.05***	**1.00**	**.29**	**3.49**	**<.05***
Determiner type	.00	.03	.08	.94	.13	.29	.44	.66
Relatedness	.00	.03	−.09	.93	−.01	.28	−.03	.98
Skipping Pre-target × Determiner Type	−.04	.05	−.86	.39	−.01	.40	−.02	.98
Skipping Pre-target × Relatedness	−.02	.05	−.51	.61	−.19	.41	−.48	.63
Determiner Type × Relatedness	.00	.04	.12	.90	.18	.37	.51	.61
Skipping Pre-target × Determiner Type × Relatedness	.02	.07	.34	.74	.79	.57	1.39	.16
	First pass				Probability of skipping			
Length pre-target	−.01	.01	−1.45	.15	**−.10**	**.05**	**−2.06**	**<.05***
Length target	.00	.01	.43	.67	**−.27**	**.04**	**−7.28**	**<.05***
Frequency target	**−.07**	**.02**	**−4.31**	**<.05***	**.36**	**.09**	**4.19**	**<.05***
Skipping pre-target	.06	.04	1.34	.18	**−1.23**	**.32**	**−3.83**	**<.05***
Determiner type	−.01	.04	−.20	.85	.10	.23	.44	.66
Relatedness	−.04	.04	−1.03	.30	.27	.22	1.25	.21
Skipping Pre-target × Determiner Type	−.05	.06	−.74	.46	.02	.45	.05	.96
Skipping Pre-target × Relatedness	.01	.06	.12	.91	−.15	.44	−.35	.73
Determiner Type × Relatedness	.01	.05	.19	.85	.02	.29	.07	.94
Skipping Pre-target × Determiner Type × Relatedness	.04	.09	.43	.67	.10	.62	.16	.87

under a spreading activation mechanism (Collins & Loftus, 1975; Forster, 1976; Levelt et al., 1999; Morton, 1979). Consistent with this account, we found facilitated processing of the target word in single fixation and total reading times, and similar patterns were observed in first fixation and first pass times. Crucially, semantic facilitation was not observed in the determiner conditions. Thus, the source of the facilitation effect 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 post-target regions. One explanation for the contrast between German and English is that the semantic relationship between the antecedent and the target word was weaker in English than in German, as shown by the relatedness norming task, where the difference between the related and unrelated pairs was 2.8 in English and 3.71 in German. We think that this explanation is unlikely because the antecedent and target words in English were judged as strongly related (on average 5.62 out of 7 points), although to a smaller extent than in German. Therefore, our ratings suggest that the semantic relationships in English were large enough that reactivation-based facilitation should have been observable in the eye-tracking experiment if it were indeed there. However, we cannot conclusively rule out that the lack of an effect specific to pronouns in English would have been observed with even more strongly related antecedent–target pairs. More research will be needed to address this possibility.

The current study assessed the existence of semantic relatedness effects in German and English separately, instead of comparing the two languages directly. This was done to avoid comparing between experiments that differed in their participants and that used different antecedent–target word pairs, which could have confounded the strength of the semantic relations and other parameters across languages. By using within-language determiner controls, we could ensure that lexical differences and strength of association were identical for the experimental and control conditions (pronouns and determiners respectively). However, to provide statistical support for the contrasting profiles observed between German and English in the semantic conditions, we conducted an additional analysis on the relatedness difference scores in the pronoun con-

Table 15
Linear Mixed Effects Model Estimates of Logged Reading Times in Experiment 2 (English) in the Phonological Conditions at the Post-target Region

Post-target region	$\hat{\beta}$	*SE*	*t/z*	*p*	$\hat{\beta}$	*SE*	*t/z*	*p*
	Single fixation				Total time			
Determiner type	.03	.03	1.22	.23	−.05	.03	−1.80	.08
Relatedness	.02	.03	.81	.42	−.01	.03	−.20	.84
Determiner × Relatedness	.00	.05	.01	.99	−.02	.05	−.36	.72
	First fixation				Probability of regression			
Determiner type	.03	.02	1.64	.10	−.28	.17	−1.65	.10
Relatedness	.02	.02	1.09	.28	−.19	.15	−1.23	.22
Determiner × Relatedness	−.02	.04	−.66	.51	−.02	.28	−.09	.93
	First pass				Probability of skipping			
Determiner type	.01	.02	.42	.68	.10	.15	.65	.51
Relatedness	.03	.02	1.11	.27	.21	.14	1.46	.14
Determiner × Relatedness	−.04	.04	−.96	.34	.01	.24	.05	.96

ditions of Experiment 1 (German) and Experiment 2 (English). The by-subject difference scores ($n = 60$ per language) at the target region in each of the four measures of interest (single fixation, first fixation, first pass, and total reading times) were compared using a nonparametric test, the Wilcoxon signed-rank test (Bauer, 1972) with language (English vs. German) as a predictor. Our goal was to assess whether the difference scores in the pronoun conditions were significantly larger in German than in English. This was the case in single fixation ($p < .05$) and first pass ($p < .05$). The difference was marginal in first fixation ($p = .09$). In total time, there was not a significant difference between the German and English difference scores ($p = .16$), consistent with the claim that German and English participants both showed relatedness effects in late measures.

We propose that the difference in the semantic conditions across English and German 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 et al., 1999), English speakers might not reactivate the antecedent's lemma upon encountering a pronoun, which results in the lack of spreading activation to semantically related words.

In English, semantic facilitation affected the pronoun and determiner conditions in late reading measures at the post-target region. We suggest that these 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.[2]

We propose that in languages with syntactic gender, licensing a pronoun requires the retrieval of the grammatical gender of its antecedent noun in the lexicon. The gender of inanimate nouns such as *das Bild* (*the painting*) is only grammatical and will be stored in the lexicon. The gender of animate nouns, such as *der Zeichenlehrer* (*the drawing teacher*) is both conceptual and grammatical, such that pronouns will reactivate both the lexical entry and discourse representation of their antecedent noun. As grammatical gender is stored as part of a word's lemma, which includes both syntactic and semantic properties (Levelt et al., 1999), when the pronoun's antecedent is reactivated, speakers of languages with syntactic gender should have joint reaccess to its syntactic and semantic features. In contrast, phonological properties are stored at a different level, belonging to the word's lexeme, such that lemma and lexeme reactivation can occur independently and with different time courses.

Interestingly, our English eye-tracking findings differ from previous cross-modal lexical decision experiments, which did find semantic facilitation effects (Leiman, 1982; 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

[2] 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's*] . . . ") an alternative account 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.

a strategy to focus on semantic antecedent information to improve their performance in the task, as opposed 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 (1982), lexical decisions to words after pronouns took on average 781 ms (unrelated: 824 ms; 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 event-related potential measures using a sentence comprehension paradigm.

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 operating at these levels of representation might differ and might yield qualitative differences in information retrieval during comprehension.

Phonological Effects

German comprehenders showed 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 but that 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 at the target region suggests that pronouns might initially 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. For instance, Levelt's model of production (e.g. Levelt et al., 1999) involves a two-stage sequential architecture, where activation of a word's lexeme occurs after the activation of its lemma. Thus, the reactivation of the antecedent's form might have taken place after readers had already moved their gaze past the target word. At the post-target region, readers likely shifted their processing to the post-target word, which was not systematically related to the antecedent noun, and thus, was unlikely to elicit phonological interference. In other words, if inhibition effects are due to coactivation, they might not obtain if the phonology of the antecedent was activated too late, after readers' eyes had already moved to the following word. However, this account would seem to predict that delayed pronoun-specific phonological effects should have been observed in the post-target region, a pattern that was not observed.

Conclusion

We used eye movements during 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.

References

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

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. http://dx.doi.org/10.1016/j.jml.2007.12.005

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39,* 445–459.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. Retrieved from http://CRAN.R-project.org/package=lme4

Bauer, D. F. (1972). Constructing confidence sets using rank statistics. *Journal of the American Statistical Association, 67,* 687–690. http://dx.doi.org/10.1080/01621459.1972.10481279

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990.

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. http://dx.doi.org/10.1006/jmla.1997.2528

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 and Language, 56,* 103–128. http://dx.doi.org/10.1016/j.jml.2006.07.005

Carreiras, M., Garnham, A., & Oakhill, J. (1993). The use of superficial and meaning-based representations in interpreting pronouns. Evidence from Spanish. *The European Journal of Cognitive Psychology, 5,* 93–116. http://dx.doi.org/10.1080/09541449308406516

Carroll, P., & Slowiaczek, M. L. (1986). Constraints on semantic priming in reading: A fixation time analysis. *Memory & Cognition, 14,* 509–522. http://dx.doi.org/10.3758/BF03202522

Chow, W. Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology, 5,* 630. http://dx.doi.org/10.3389/fpsyg.2014.00630

Cloitre, M., & Bever, T. G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language and Cognitive Processes, 3,* 293–322. http://dx.doi.org/10.1080/01690968808402092

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–428. http://dx.doi.org/10.1037/0033-295X.82.6.407

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, NY: Academic Press.

Ehrlich, K., & Rayner, K. (1983). Pronoun Assignment and Semantic Integration during Reading: Eye Movements and Immediacy of Processing. *Journal of Verbal Learning and Verbal Behavior, 22,* 75–87.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam, the Netherlands: 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, UK: 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. http://dx.doi.org/10.1006/jmla.1995.1003

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. http://dx.doi.org/10.1006/jmla.1994.1003

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language, 42,* 526–544.

Heim, I. (1988). *The semantics of definite and indefinite noun phrases*. New York, NY: Garland Publishers.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86,* 157–171. http://dx.doi.org/10.1016/j.jml.2015.10.007

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446.

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. http://dx.doi.org/10.1037/0278-7393.27.4.1058

Kempen, G., & Huijbers, P. (1983). The lexicalisation process in sentence production and naming: Indirect election of words. *Cognition, 14,* 185–209. http://dx.doi.org/10.1016/0010-0277(83)90029-X

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2014). *lmerTest: Tests for random and fixed effects for linear mixed effect models*. Retrieved from http://CRAN.R-project.org/package=lmerTest

Leiman, J. M. (1982). *A chronometric analysis of referent assignment to pronouns*. (Unpublished doctoral dissertation). Detroit, MI: Wayne State University.

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,* 1–38. http://dx.doi.org/10.1017/S0140525X99001776

Lucas, M. M., Tanenhaus, M. K., & Carlson, G. N. (1990). Levels of representation in the interpretation of anaphoric reference and instrument inference. *Memory & Cognition, 18,* 611–631. http://dx.doi.org/10.3758/BF03197104

McKoon, G., & Ratcliff, R. (1981). The comprehension processes and memory structures involved in instrumental inference. *Journal of Verbal Learning and Verbal Behavior, 20,* 671–682. http://dx.doi.org/10.1016/S0022-5371(81)90238-3

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. http://dx.doi.org/10.1037/0278-7393.20.5.1219

Meyer, A. S., & Bock, K. (1999). Representations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory and Language, 41,* 281–230. http://dx.doi.org/10.1006/jmla.1999.2649

Morris, R. K., & Folk, J. R. (1998). Focus as a contextual priming mechanism in reading. *Memory & Cognition, 26,* 1313–1322. http://dx.doi.org/10.3758/BF03201203

Morton, J. (1979). "Word recognition." In J. Morton & J. C. Marshall (Eds.), *Psycholinguistics Series: Structures and processes* (Vol. 2). London, UK: 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, NJ: Lawrence Erlbaum.

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

Nicol, J. L., Fodor, J. D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1229–1238. http://dx.doi.org/10.1037/0278-7393.20.5.1229

Paterson, K., Alcock, A., & Liversedge, S. P. (2011). Morphological priming during reading: Evidence from eye movements. *Language and Cognitive Processes, 26,* 600–623. http://dx.doi.org/10.1080/01690965.2010.485392

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. http://dx.doi.org/10.3758/PBR.16.1.43

R Core Team. (2016). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Schmitt, B. M., Meyer, A. S., & Levelt, W. J. M. (1999). Lexical access in the production of pronouns. *Cognition, 69,* 313–335. http://dx.doi.org/10.1016/S0010-0277(98)00073-0

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception & Psychophysics, 74,* 5–35. http://dx.doi.org/10.3758/s13414-011-0219-2

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. http://dx.doi.org/10.1016/S0010-0277(03)00161-6

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. http://dx.doi.org/10.1037/0278-7393.23.4.787

Received August 10, 2015
Revision received August 27, 2016
Accepted August 31, 2016 ■