

Processing mismatching gendered possessive pronouns in L1 Dutch and L2 French

Gert-Jan Schoenmakers¹, Theresa Redl¹, Sebastian Collin¹, Rozanne Versendaal², Peter de Swart¹ and Helen de Hoop¹

¹Radboud University | ²Utrecht University

Abstract The results of a self-paced reading experiment show that reading times in Dutch increase when there is a gender mismatch between the subject and a subsequent possessive pronoun, signaling an increase in processing difficulty. We hypothesized that Dutch learners of French incorrectly apply the rules of their L1 in their L2 and should therefore also show an increase in reading times in French upon encountering a possessive pronoun for which grammatical gender differs from the biological gender of the subject (the possessor). At the same time, we expected that they would have no or less difficulties in processing ungrammatical French sentences in which the biological gender of the subject/possessor matches the gender of the possessive pronoun. We did not find either of these effects in a second self-paced reading experiment. We assume that the Dutch learners of French parse the foreign language sentences in a shallow fashion.

Keywords self-paced reading, gender features, syntactic agreement, semantic agreement, shallow processing

Article history

Received: June 2, 2021

Accepted: March 28, 2022

Online: July 14, 2022

Corresponding author

Gert-Jan Schoenmakers,
gert-jan.schoenmakers@ru.nl

Copyright

© Author(s); licensed under Creative Commons Attribution 4.0. This allows for unrestricted use, as long as the author(s) and source are credited.

Funding information

None

Conflicting interests

The authors have declared that conflicts of interest did not exist.

1 Introduction

Dutch students of French in secondary education are known to struggle with the gender features of its third person possessive pronouns, because these pronouns agree syntactically with the head noun (the possessee), whereas the gender of the possessive pronoun in their L1 agrees semantically with the gender of the possessor. In Dutch, the gender of the possessive pronouns *zijn* 'his' and *haar* 'her' refers to the gender of the pronoun's referent, i.e., the possessor. In French, by contrast, the semantic gender of the possessor does not determine the choice for the grammatically masculine pronoun *son* 'his/her' or the grammatically feminine pronoun *sa* 'his/her'. The gender feature in French is instead dependent on and agrees with the grammatical gender of the possessee, i.e., the head noun that the determiner combines with.¹ Hence, the gender of the third person singular possessive pronoun in French is an instantiation of *syntactic agreement* (Corbett, 1979),

while it is not a grammatical property of the possessive pronoun in Dutch – nor in English (Slevc et al., 2007).

With regard to the difference between the gender systems of Dutch and French pronouns, this article addresses two questions. First, we seek to determine whether gender mismatches between the subject and a subsequent possessive pronoun, as in (1), lead to processing difficulties in Dutch.

- (1) *Caroline heeft zijn piano in Londen gekocht.*
‘Caroline bought his piano in London.’

Both in Dutch and English, sentences like (1) are pragmatically odd in isolation. The possessive pronoun cannot anaphorically refer to the subject because of a gender mismatch, hence (1) raises the question *whose* piano Caroline bought in London. We expect this to lead to extra processing costs. Piepers and Redl (2018) found that speakers of Dutch consider sentences with such a gender mismatch, for example *Judith heeft zijn outfit samengesteld* ‘Judith put together his outfit’ significantly less natural than sentences in which the subject’s gender matches with the gender of the possessive pronoun.

The second question addressed in this study is whether the difference in type of agreement between Dutch and French possessive pronouns leads to processing difficulties for Dutch learners of French when there is a gender mismatch between the semantic gender of the subject and the syntactic gender of the possessive pronoun, as in (2).

- (2) *Caroline a acheté son piano à Londres.*
‘Caroline bought her(M) piano in London.’

If this is the case, they might have no or less processing difficulties when reading ungrammatical French sentences with matching biological and grammatical gender, as in (3).

- (3) **Caroline a acheté sa piano à Londres.*
‘Caroline bought her(*F) piano in London.’

Section 2 presents an overview of relevant literature on L1 and L2 processing of gendered pronouns. Section 3 reports on a self-paced reading experiment that tested the processing of sentences such as (1). If indeed a processing effect of gender mismatches is found in Dutch, we hypothesize that Dutch learners of French may show similar processing difficulties in reading grammatical French sentences such as (2), and less or no difficulties in reading ungrammatical sentences like those in (3), i.e. show a transfer effect from their L1. These predictions are tested in Section 4. Section 5 discusses our findings in light of the previous literature, and Section 6 presents our conclusions.

2 Previous literature on processing gendered pronouns

Gender mismatches have been found to hamper processing of the personal pronouns *he* and *she* in L1 English. For instance, Carreiras et al. (1996) conducted a self-paced reading task using experimental items with stereotypically male or female participants (e.g., *doctor, nurse*) and found a significant slowdown when a later pronoun did not match the stereotypical gender of its antecedent. Nieuwland (2014) conducted an EEG study and found clear effects of pronouns that did not match the gender of the only available antecedent, as in *The boy thought that she would win the race*. Further, Dong et al. (2015) conducted two self-paced reading experiments with highly proficient Mandarin Chinese learners of English, and found a gender-mismatch effect on the pronoun for sequences like *Mark goes to the zoo to watch animals every day after work for a good rest. She considers it the best way to relax and maintain a good mood*, but only when the gender of the antecedent subject was enhanced by a gender-consistent picture, in this case, a picture of a man.

As for possessive pronouns, a mismatch between the biological gender of the possessor and the biological gender of the possessee, as in *his mother* or *her father*, can also lead to errors in the production of *his* and *her* in English. Slevc et al. (2007) found in an experimental study that native speakers of English made more errors in producing the possessive pronoun when there was a mismatch between the genders of the possessor and the possessee within the noun phrase (5.1% errors like *Victor gave her sister a present*) than when there was a match between the genders (1.7% errors like *Victoria gave his sister a present*). This error rate was independent of the lexical noun used, as long as the gender of the referent was known. Thus, the noun *cousin*, which can refer to both a female and a male, would lead to the same amount of gender errors as *sister* when the context had made it clear that *cousin* referred to a woman. Slevc et al. (2007) conclude that the gender mismatch is only a mismatch in the genders of biological referents in English, and not due to lexical or syntactic properties of the words used.

Possessive pronouns in Romance languages do not have semantic gender reference, but show syntactic gender agreement. Antón-Méndez (2011) found in an error elicitation experiment that Spanish L2 speakers of English made significantly more errors than Dutch speakers when there was a gender mismatch between the genders of the possessor and the possessee within the noun phrase, such as in *her father*. However, they only found this result for human possessee, not when the head was an inanimate noun that would have a syntactically mismatching gender in Spanish, but not in English. This may suggest that the errors Romance learners of English make in producing the correctly gendered possessive pronoun is not caused by the fact that they have syntactic gender agreement but rather by the lack of semantic agreement in their native language. Supporting evidence for this comes from a study with Mandarin Chinese learners of English, who make similar errors in the production of *his* or *her* when there is a gender mismatch between the possessor and the possessee within the noun phrase (Pozzan & Antón-Méndez, 2017),

despite the fact that Mandarin Chinese does not have syntactic gender agreement. The reason might be that Mandarin Chinese does not make a distinction between masculine and feminine pronouns in spoken language, which suggests that the increased number of errors in the production of gendered possessive pronouns in English must be the result of the lack of semantic gender agreement in one's native language.

Native speakers of English (Slevc et al., 2007) as well as learners of English with different native backgrounds (Antón-Méndez, 2011; Pozzan & Antón-Méndez, 2017; see also White et al., 2007) thus make errors when they have to produce the possessive pronouns *his* or *her* within a noun phrase that has a mismatch between the semantic genders of the possessor and the possessee. These errors in production are not mirrored in comprehension (Pozzan & Antón-Méndez, 2017). Although Pozzan and Antón-Méndez (2017) found more production errors for Mandarin Chinese learners of English in the case of mismatching genders within the noun phrase, they did not find evidence for processing difficulties in a concomitant comprehension task. The experimental items prompted participants to place an object next to some other object or character (e.g., *Give the apple to his little sister*) and here the possessive pronouns were correctly interpreted, independently of the gender mismatch between the pronoun and the head noun's referent, e.g., *sister* and *his*.

In a speeded acceptability judgment experiment, Lago et al. (2019) tested whether English and Spanish learners of German differed in their processing of third person possessive pronouns, which in German agree both semantically with the possessor (*sein* 'his' versus *ihr* 'her') and syntactically with the possessee. Whereas possessive pronouns in English (like in Dutch) only semantically agree with the possessor, in Spanish (like in French) they only syntactically agree in gender with the possessee. Participants were asked to judge the acceptability of German sentences with a possessive construction while their reaction times were being recorded. The stimulus sentences were manipulated such that the referent of the possessive pronoun (the possessor) did or did not agree in semantic gender with the subject of the sentence, as illustrated in (4) (Lago et al., 2019, p. 325).

- | | | |
|-----|--|------------------------|
| (4) | a. <i>Frau Schmidt küsste ihre Mutter.</i>
'Ms. Schmidt kissed her(F) mother.' | felicitous; match |
| | b. <i>Herr Schmidt küsste seine Mutter.</i>
'Mr. Schmidt kissed his(F) mother.' | felicitous; mismatch |
| | c. # <i>Frau Schmidt küsste seine Mutter.</i>
'Ms. Schmidt kissed his(F) mother.' | infelicitous; match |
| | d. # <i>Herr Schmidt küsste ihre Mutter.</i>
'Mr. Schmidt kissed her(F) mother.' | infelicitous; mismatch |

The sentences in (4a) and (4b) are felicitous, because the subject and the possessive pronoun have the same gender and can therefore be interpreted as coreferential. The

sentences in (4c) and (4d), by contrast, are pragmatically infelicitous, marked by the sign #, because there is a mismatch between the semantic genders of the subject and the possessive pronoun, which means that the subject and the possessor cannot be coreferential. These sentences thus correspond to the Dutch sentence in (1) above, which implies that a new discourse referent is introduced by the possessive pronoun in (4c) and (4d).

An additional manipulation in Lago et al.'s (2019) experiment is what they call *gender match* and *gender mismatch*. This is quite confusing, because 'felicitous' and 'infelicitous' sentences can also be described as involving a gender match or mismatch. What the authors mean is that the gender of the subject *Frau Schmidt* 'Ms Schmidt' corresponds to the gender of the object *Mutter* 'mother' in the matching sentences (4a) and (4c), while there is a mismatch between the genders of the subject *Herr Schmidt* 'Mr Schmidt' and the object *Mutter* 'mother' in (4b) and (4d). Note that none of the sentences in (4) are ungrammatical, because the grammatical gender of the possessive pronoun agrees with the gender of the head noun *mother* in all cases.

While the accuracy in judgments was similar across the participant groups, the Spanish participants took longer to respond to the infelicitous sentences in (4c) and (4d) than the English participants, suggesting that Spanish learners of German, whose native language lacks the semantic gender distinction between *ihr* 'her' and *sein* 'his', had more difficulties in judging the acceptability of these sentences than the English learners of German.

Lago et al. (2019, p. 333) did not find a difference between their 'gender matching' and 'gender mismatching' conditions, and state that "[t]his was unexpected, as previous production studies had found that possessive errors were more frequent when the possessor and possessee noun mismatched in gender". They had expected Spanish learners to make more errors in judging the acceptability of the 'mismatching' sentences in (4b) and (4d) than the English learners. However, their 'matching' condition does not in fact pertain to the gender agreement between the possessor and the possessee within the noun phrase, as in the previous studies discussed, but rather to the gender agreement between the subject and the object. Hence, they code the infelicitous sentence in (4c) as a gender match and the infelicitous sentence in (4d) as a gender mismatch, but when considering the gender agreement between the possessor and possessee within the noun phrase, this should be the other way around. A gender mismatch between the male possessor (*sein* 'his') and the female possessee (*Mutter* 'mother') occurs in (4b) and (4c), but not in (4d). Coreferentiality between the subject and the possessor is not possible in the infelicitous sentences (4c) and (4d). Therefore, there is no gender mismatch between the female possessor *ihre* 'her' and the female possessee *Mutter* 'mother' in (4d), but only between the subject *Herr Schmidt* 'Mr. Schmidt' and the object *Mutter* 'mother'.

Thus, while Pozzan and Antón-Mendez (2017) did not find processing difficulties for Mandarin Chinese speakers of English in case of a gender mismatch between the possessor and the possessee (e.g., *Give the apple to his little sister*), Lago et al. (2019) did

not in fact test the effect of such a gender mismatch. They instead gauged the impact of a gender mismatch between the subject and the object, for which they did not find evidence. Note that processing difficulties due to such a subject-object gender mismatch were not expected on the basis of previous research.

In a subsequent implicit self-paced reading experiment, Lago et al. (2019) found that Spanish natives were less affected by the infelicitous possessive pronouns in sentences such as (4c) and (4d) than English natives, indicating that not only their production of possessive pronouns in German is affected, but also their reading comprehension (*pace* Pozzan & Antón-Méndez, 2017). Again, it may just be the lack of semantic gender agreement in their L1 that causes these difficulties for Spanish learners.

Unlike German, Dutch does not have syntactic gender agreement between the possessive pronoun and the head noun. Extrapolating the results of the previous studies on the processing of gender mismatches in pronouns (Carreiras et al., 1996; Dong et al., 2015; Nieuwland, 2014; Slevc et al., 2007) to the present study, we can expect native speakers of Dutch to be hampered by a gender mismatch in Dutch sentences such as (1) above. If so, Dutch learners of French may also experience processing difficulties when reading sentences like (2), in which the syntactic gender of the possessive pronoun does not match the semantic gender of the subject. Versendaal (2013) conducted a self-paced reading study to test this with three groups of Dutch students of French: high school students in their third year, high school students in their sixth year, and first year university students (and additionally a control group of French native speakers). She found a sharp decrease in reading speed between the sixth year high school students and the university students, as well as between the university students and the native speakers. However, for none of the groups a significant difference was found between the reading times of matching sentences, such as *Marie a vendu sa maison* 'Marie sold her(F) house', and mismatching sentences, as in *Jean a vendu sa maison* 'Jean sold his(F) house'. Versendaal suggests that Dutch learners do not at all pay attention to the syntactic or semantic gender information when reading a French sentence. She links this to the idea of *good enough processing* (Ferreira et al., 2002), which submits that L2 learners are only concerned with understanding the sentences, and not with their morphosyntactic characteristics.

However, Versendaal (2013) only tested grammatical French sentences, which makes it impossible to conclude that the Dutch students did not construct a grammatical representation of the French sentences they read. Moreover, she did not check whether Dutch students show an increase in reading times with semantically mismatching sentences in their native language. If they do not show an effect of mismatches in Dutch sentences such as (1) above, we would not expect them to show an effect of mismatches between syntactic and semantic gender agreement in French sentences such as (2). Therefore, following up on Versendaal's preliminary results, we conducted two self-paced reading experiments among native Dutch students in secondary education, on which we report next.

3 Dutch experiment

The hypothesis is investigated whether native speakers of Dutch have longer reading times in processing Dutch sentences such as (1) above, in which there is a mismatch between the semantic genders of the subject and the subsequent possessive pronoun, in comparison to sentences without such a mismatch.

3.1 Participants

A total of 75 participants at two high schools in the Netherlands volunteered. Two participant groups were included in the experiment to establish a degree of similarity between Experiment 1 and Experiment 2: 44 participants were in their 3rd year of vwo (the pre-academic track), approximately 14 years old, and 31 participants were in their 5th or 6th year, approximately 16 to 18 years old. All participants were tested in Dutch. Consent was obtained in two ways. First, after the schools showed initial interest in participating in the experiment, an information document describing the procedure and data storage was sent to school officials and signed by them. Second, parents received a letter providing similar information. Parents were given the opportunity to withdraw consent and exclude their child from the experiment by informing the school. No parent made use of this option. The experiments reported in this paper were assessed and approved by the Ethics Assessment Committee Humanities of Radboud University (ETC-1438).

3.2 Materials and design

The experiment employed a 2×2 design. The two-level factor *Education* varied between participants, featuring vwo 3rd graders and 5th/6th graders, the latter being subsumed in one group. We moreover varied the two-level factor *Congruency* within participants. The congruent condition featured items in which the gender of the subject matched the gender of the possessive pronoun (5a and 5b); the incongruent condition featured items with a gender mismatch (5c and 5d). The experiment contained 20 stimulus items, with a proper name (half male, half female) followed by an auxiliary, a possessive pronoun (half masculine, half feminine), a noun, a prepositional phrase, and a past participle. All items were grammatical. Items were pseudo-randomized to avoid clustering of items in the same condition.

- | | | |
|-----|--|-------------|
| (5) | a. <i>Christel heeft haar croissant in het park opgegeten.</i> | congruent |
| | b. <i>Lucas heeft zijn croissant in het park opgegeten.</i> | congruent |
| | c. <i>Christel heeft zijn croissant in het park opgegeten.</i> | incongruent |
| | d. <i>Lucas heeft haar croissant in het park opgegeten.</i> | incongruent |
| | ‘Christel/Lucas ate her/his croissant in the park.’ | |

In addition, 40 unrelated filler items and 24 control statements were added to the experiment. The control statements were added to verify the participants' attention and to test their semantic processing of the experimental items. These control statements could be judged as 'correct' or 'incorrect' with minimal logical reasoning (half of them were correct). To give one example, the sentence *Felix heeft zijn broek in de kast gezocht* 'Felix looked for his pants in the closet' was followed by the control statement *Felix is een broek kwijt* 'Felix has lost a pair of pants', with the intended answer 'correct'.

3.3 Procedure

The experiment was programmed and run in OpenSesame and took the form of a moving window self-paced reading task. Participants were tested in small groups in a computer lab at their school. They first read the instructions for the self-paced reading task. They saw seven practice items and were given the opportunity to ask clarification questions. They were then presented with the experimental items. Participants were allowed to abort the experiment at any given moment.

3.4 Analysis

We first checked the average accuracy with which each control statement was answered. We eliminated control statements that were answered correctly in less than 80% of the trials before excluding participants based on their responses. This was the case for two control items, which were removed from further analysis. We subsequently removed all data from participants who responded correctly to less than 80% of the remaining control statements: one participant, who scored 63.6% on the control statements, had to be removed from further analysis. This left us with data from 74 participants (43 in 3rd grade of vwo).

We analyzed the reading times of three regions of interest: (i) the possessive pronoun, (ii) the possessee, and (iii) the preposition; see (6). The crucial measure is the reading time for words in the first region, as the gender (mis)match happens at this point in Dutch. That is, the possessive pronoun agrees with the subject in Dutch, i.e. *haar* 'her' agrees with *Christel* in (6). Given that processing may not be complete at the target word, however, the effect may *spill over* in the subsequent regions (Mitchell, 1984). The two words further downstream (the possessee noun and the preposition) are therefore also included in the analysis as well.

- (6) *Christel heeft [haar]₁ [croissant]₂ [in]₃ het park opgegeten.*
 'Christel ate her croissant in the park.'

In line with Baayen and Milin (2010), we removed physically impossible reading times, defined for our experiment as 50ms, and extremely high values defined as 5000ms, which

are no longer thought to reflect processing in reading per se. Four observations were above the 5000ms threshold and one observation was below the 50ms threshold. These observations were removed from further analysis. We applied a reciprocal transformation ($1/x$) on the remaining data for each region of interest to account for the skew inherent to a distribution of reading times (see Kliegl et al., 2010). We removed outlying data points using a standard deviation of 2.5 as a threshold. Standard deviations were determined on the distribution of each individual condition and for each individual region of interest. We removed 22 observations in region 1 (1.5%), 28 observations in region 2 (1.9%), and 29 observations in region 3 (2%).

The remaining data were analyzed using linear mixed effects models in R (R Core Team, 2017) using the lme4 package (Bates et al., 2015b). Following Barr et al. (2013), we used maximal random effect structures whenever possible. In case of non-convergence or signs of overparameterization, we reduced the complexity of random effect structures by following the following steps: 1) disabling random correlations, 2) removing higher order effects from the random component, and 3) using a random intercept only. Models were checked for overparameterization by applying Principal Component Analysis (PCA) on the covariance matrices of random effect estimates using the RePsychLing package (Bates et al., 2015a). *Congruency* and *Education* as well as the interaction between the two served as fixed effects across all regions. Sum contrasts were used. *Congruency* was coded as 1 for congruent and -1 for incongruent; *Education* was coded as 1 for 3rd graders and -1 for 5th/6th graders. *P*-values were obtained using the normal approximation to the *t*-statistic.

3.5 Results

3.5.1 Region of interest 1 – possessive pronoun

A linear mixed effects model was fit to the reciprocal reading times on the possessive pronoun. Random intercepts per participant were included, but no random slopes were fitted per participant. For items, random intercepts and random slopes for *Education* per item were fitted. Correlation parameters were suppressed. There were no significant effects; the transformed data on which the analysis was performed are visually represented in Figure 1. Note that higher numbers and bars actually represent lower reading times prior to the transformation because of the reciprocal transformation ($1/x$).

3.5.2 Region of interest 2 – noun

Another linear mixed effects model was fit to the reciprocal reading times on the possessee noun. The full random effects structure permitted by the design was included, but no correlation parameters were estimated. There was a significant effect of *Congruency* ($\beta = -0.000074$, $SE = 0.000019$, $t = -3.892$, $p < 0.001$). Figure 2 visually represents the transformed data and shows that 3rd and 5th/6th graders read the possessee noun faster in the congruent conditions (5a,b) than in the incongruent conditions (5c,d). Figures of

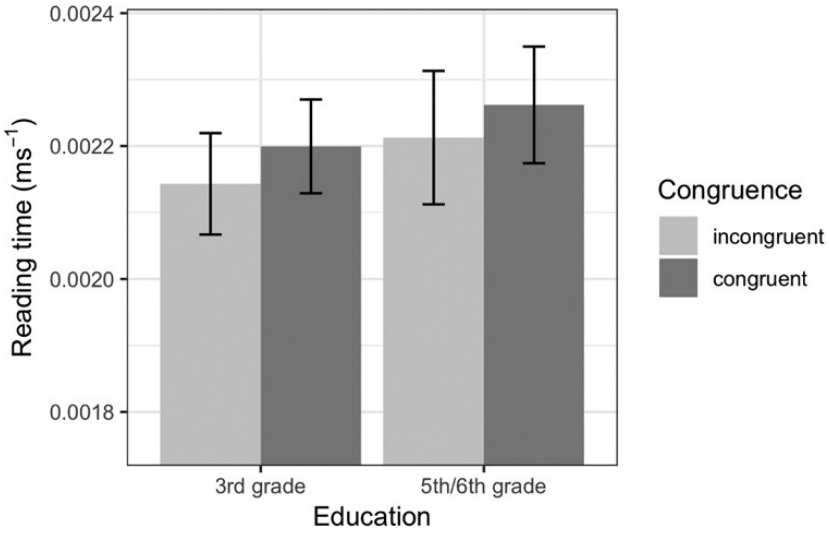


Figure 1 Reading times in reciprocal milliseconds for region 1 (possessive pronoun) for the Dutch experiment with 95% confidence intervals. Note that higher reciprocal reading times correspond to lower reading times prior to transformation.

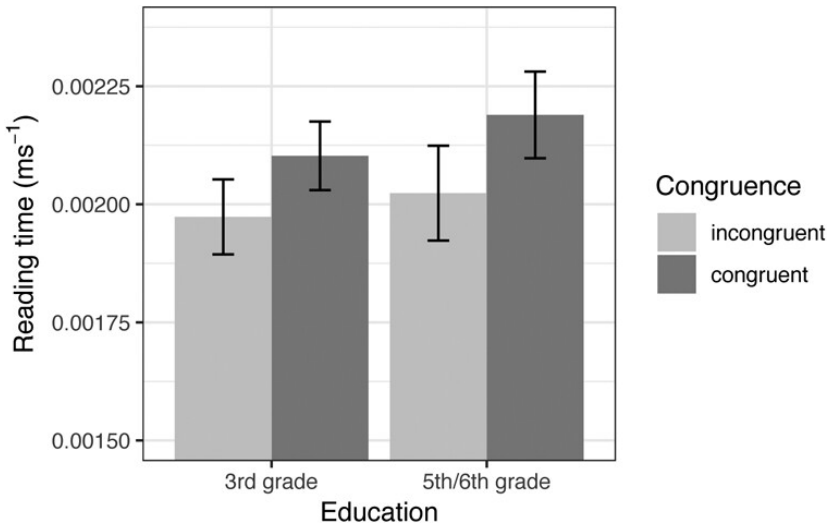


Figure 2 Reading times in reciprocal milliseconds for region 2 (noun) for the Dutch experiment with 95% confidence intervals. Note that higher reciprocal reading times correspond to lower reading times prior to transformation.

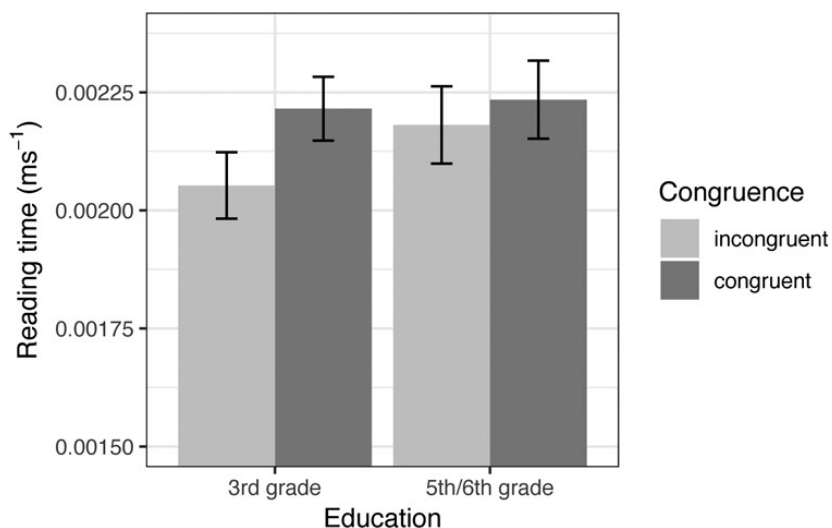


Figure 3 Reading times in reciprocal milliseconds for region 3 (preposition) for the Dutch experiment with 95% confidence intervals. Note that higher reciprocal reading times correspond to lower reading times prior to transformation.

the untransformed data for regions 2 and 3 and the mean reading times and standard deviations per level can be found in the Supplementary material.

3.5.3 Region of interest 3 - spillover region

The final model was fit to the reciprocal reading times on the preposition, i.e. the spillover region, and included by-participant random intercepts, but no by-participant random slopes. By-item random intercepts were included, as well as by-item random slopes for *Congruency* and *Education* (but not the interaction between the two). There was a significant main effect of *Congruency* ($\beta = -0.000052$, $SE = 0.000017$, $t = -3.165$, $p = 0.002$). Figure 3 displays the transformed data and shows that 3rd and 5th/6th graders read the preposition following the possessee noun in the congruent conditions (5a,b) faster than in the incongruent conditions (5c,d).

3.5.4 Discussion

The results show that reading times are increased for possessive pronouns that are not congruent with the subject, as observed in the two regions following the possessive pronoun. This indicates that native speakers of Dutch experience processing difficulties when they read pragmatically infelicitous sentences in which the subject and possessive pronoun do not agree in gender, such as in (5c,d) above. This finding is in line with Piepers and Redl's (2018) judgment data, as well as with the experimental literature on similar structures in English, where a recurrent finding is that processing is hindered

by mismatches in gender between a potential antecedent and a subsequent pronoun (Carreiras et al., 1996; Dong et al., 2015; Nieuwland, 2014; Slevc et al., 2007).

The next section investigates whether this congruency effect carries over to the processing of French sentences by Dutch learners. If so, Dutch learners of French are expected to experience processing difficulty when they encounter a gender mismatch between the subject's semantic gender and the possessive pronoun's syntactic gender, as in (2) above, but no or less difficulty when the genders match in an ungrammatical sentence such as (3) above. Finally, we expect that the severity of the congruency effect decreases the more proficient the learners are in French, whereas their sensitivity to the ungrammatical structures increases. The control group, which consists of native speakers of French, is expected to be affected by grammaticality only, and not by the gender congruency between subject and possessive pronoun.

4 French experiment

The hypothesis is tested whether Dutch learners of French take longer to read French sentences such as (2) above, and if so, whether the increase in reading times diminishes or disappears when the sentence is ungrammatical, as in (3). Further, we investigate whether the putative effects are dependent on the non-native participants' experience with the French language.

4.1 Participants

A total of 103 participants were tested, 47 participants in their 3rd year of Dutch vwo, and 36 in their 5th or 6th year. The participants all had had French class since their first year in high school. The consent procedure was identical to that in Experiment 1. Parents were informed that they could refuse consent and exclude their child from the experiment, but no parent made use of this option. A further 20 participants were native speakers of French and students at the University of Lyon. The experiment was advertised at the university and prospective participants contacted the researcher via email to schedule a testing session. All participants were tested in French using the same task.

4.2 Materials and design

We employed a $3 \times 2 \times 2$ design. The two-level factor *Education* varied between participants, featuring Dutch vwo 3rd graders (ages approximately 14), Dutch 5th/6th graders (ages 16–18), and French university students (ages 18–25). The two-level factor *Congruency* varied within participants. The congruent condition featured items in which the gender of the subject matched the gender of the possessive pronoun (7b and 7d); the incongruent condition featured items with a gender mismatch (7a and 7c). Further, we

varied the two-level factor *Grammaticality*. The conditions in (7a) and (7b) were grammatical, but the conditions in (7c) and (7d) ungrammatical, as the possessive pronoun did not agree in gender with the head noun. All items followed the same pattern, with a proper name (half male, half female) followed by an auxiliary, past participle, possessive pronoun (half masculine, half feminine), a noun (half masculine, half feminine), and a prepositional phrase. None of the possessee nouns started with a vowel or a *mute h*.

- (7) a. *Christelle a mangé son croissant dans le restaurant.* incongruent, grammatical
 b. *Lucien a mangé son croissant dans le restaurant.* congruent, grammatical
 c. *Lucien a mangé sa croissant dans le restaurant.* incongruent, ungrammatical
 d. *Christelle a mangé sa croissant dans le restaurant.* congruent, ungrammatical
 ‘Christel/Lucas ate her/his croissant in the restaurant.’

Lexical items were selected from the textbook used in the third year of vwo, *Grandes Lignes*. Due to the limited vocabulary of 3rd graders in particular, however, we had to reuse sentence frames. Although the nouns from the noun phrase with the possessive pronoun were not repeated across items, the proper nouns, verbs, and prepositional phrases were. We therefore split the experiment in two blocks, and sentence frames were repeated in the second block. Each item was presented in the inverse condition in the second block. Thus, the sentence frame used in (7a) in the first block would occur as the sentence frame in (7c) in the second block, but with a different head noun (e.g., *baguette* instead of *croissant*). Items were pseudo-randomized to avoid clustering of items from the same condition.

4.3 Procedure

The experiment was programmed and run in OpenSesame. Participants were tested in small groups in a computer lab at their school or university. They first read the instructions, practiced with seven items, and were given the opportunity to ask clarification questions. They were then presented with 40 stimuli, 40 fillers, and 24 control statements (half of which were correct). The control statements were translated from the Dutch experiment. Afterwards, participants read the instructions for a second task, designed to ensure that participants knew the grammatical gender of the nouns used in the self-paced reading task. They were again given the opportunity to ask questions. Participants were presented with a list of the 40 relevant nouns and had to indicate whether the noun required the determiner *le* (grammatically masculine) or *la* (grammatically feminine) using a radio button. Participants were allowed to abort the experiment at any given moment.

4.4 Analysis

We first checked the average score of each control statement. We did this for the French and the Dutch participants separately. We decided to discard control statements that were answered correctly in less than 60% of the trials for the Dutch participants and less than 80% for the French participants. The reason to do this was to eliminate inherently difficult control statements before eliminating participants on the basis of their answers on control statements. The employed criteria differed for the Dutch and the French participants as the French participants were native speakers of the test language, while the Dutch participants were (early) learners. This led to the exclusion of twelve of the 24 control statements for the Dutch participants, and ten out of the 24 questions for the French participants. Four of the removed control statements were the same for Dutch and French participants; the rest were different. Thus, we removed all data from Dutch participants who correctly responded to less than 60% of the remaining control statements, i.e. thirteen 3rd graders and four 5th/6th graders, and all data from French participants who correctly responded to less than 80% of the remaining control statements, i.e. three participants. This left us with the data of 83 participants. As a next step, we considered whether the participants knew the gender of the words that had occurred in the experimental items. Looking at each participant individually, we removed data points from items that featured a noun of which the participant could not correctly indicate the grammatical gender. 18% of the observations were removed.

We then analyzed the reading times of the same three regions as in Experiment 1: (i) the possessive pronoun, (ii) the possessee, and (iii) the preposition; see (8). The crucial measures in this experiment are the reading time for words in the first and second region, as the gender (mis)match happens in the first region in Dutch and in the second region in French. That is, the possessive pronoun in French matches with the possessee that follows the pronoun, i.e. *son* 'his/her' agrees with *croissant* 'croissant' in (8). In Dutch, however, the possessive pronoun agrees with the possessor (cf. Experiment 1). Region 3 (the preposition) is included in the analysis as a spillover region.

- (8) Christelle a mangé [son]₁ [croissant]₂ [dans]₃ le restaurant.
'Christel ate her croissant in the restaurant.'

We removed eleven data points, as they were above 5000ms. No observation was below the 500ms threshold. We applied a reciprocal transformation ($1/x$) on the data for each region of interest to account for the skew inherent to a distribution of reading times. We then removed outlying data points, with a standard deviation of 2.5 as a threshold. Standard deviations were determined on the distribution of each individual condition and for each individual region of interest. This way, we removed 47 observations in region 1 (1.75%), 53 observations in region 2 (2%), and 52 observations in region 3 (2%).

The data were analyzed using linear mixed effects models in R (R Core Team, 2017) using the *lme4* package for modelling (Bates et al., 2015b). *Congruency* and *Education* served as fixed effects across all regions. In addition, for region 2 and region *Grammaticality* served as a fixed effect. The factor *Grammaticality* was not included in the model for the first region, as it was not relevant at this region yet; because the possessee had not been revealed yet, there was no way for the participants to know whether the agreement between the possessive pronoun and the possessee was grammatical or not. Based on our hypotheses, we further modelled the interaction between *Education* and *Congruency* as well as *Education* and *Grammaticality*. The random effects structure was determined in the same way as in Experiment 1. We used sum contrasts for the categorical variables. *Congruency* was coded as 1 for congruent and -1 for incongruent. Two contrasts were used for *Education*: one contrast compared the 3rd graders to the overall average (3rd graders = 1, 5th/6th graders = 0, French natives = -1), the other compared the 5th and 6th graders to the overall average (3rd graders = 0, 5th/6th graders = 1, French natives = -1). *P*-values were obtained using the normal approximation to the *t*-statistic.

4.5 Results

4.5.1 Region of interest 1 - possessive pronoun

A linear mixed effects model was fit to the reciprocal reading times on the possessive pronoun. The model included *Congruency* and *Education* as fixed effects, as well as the interaction effect between the two. Random intercepts were fitted on the participant and the item level. Random slopes were fitted for *Congruency* per item. No correlation parameters were estimated. There was a significant main effect of *Education* when comparing the 3rd graders to the average ($\beta = -0.00026$, $SE = 0.0005$, $t = -4.9$, $p < 0.001$) and when comparing the 5th graders to the average ($\beta = -0.00011$, $SE = 0.0005$, $t = -2.11$, $p = 0.035$), indicating that the reading times of Dutch high school students were significantly higher than those of the French university students. The transformed data on which the analysis was performed are visually represented in Figure 4. Note that higher numbers and bars actually represent lower reading times prior to the transformation because of the reciprocal transformation ($1/x$).

4.5.2 Region of interest 2 - noun

A linear mixed effects model was fit to the reciprocal reading times on the noun following the pronoun. The model included *Grammaticality*, *Education*, and *Congruency* as main effects, as well as two-way interaction effects between *Grammaticality* and *Education* on the one hand, and *Congruency* and *Education* on the other hand. Random intercepts were estimated per participant and per item. In addition, random slopes for *Congruency* and for the *Education* contrast comparing Dutch learners in the 5th/6th grade to the average were included. Correlation parameters were not estimated. There was a significant main effect of *Education* when comparing the 3rd graders to the average ($\beta = -0.00028$,

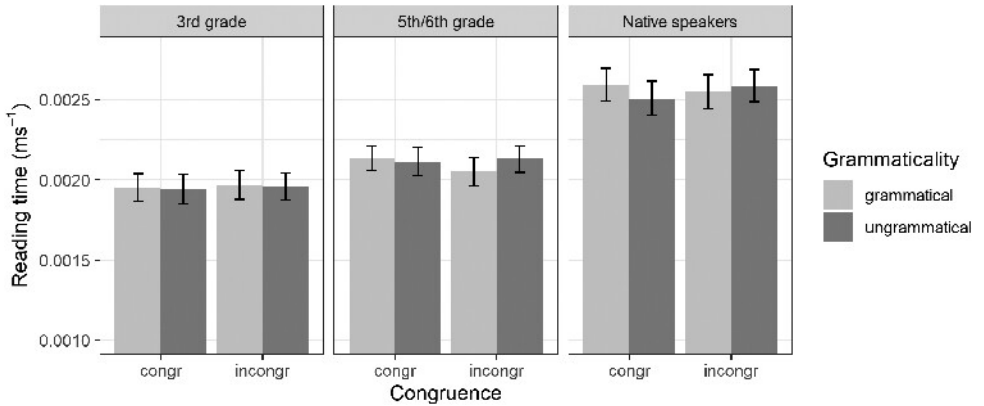


Figure 4 Reading times in reciprocal milliseconds for region 1 (possessive pronoun) for the French experiment with 95 % confidence intervals. Note that higher reciprocal reading times correspond to lower reading times prior to transformation.

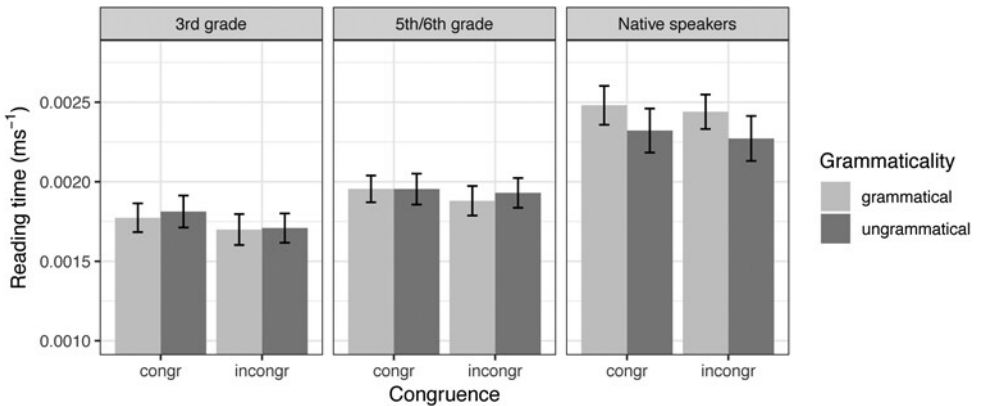


Figure 5 Reading times in reciprocal milliseconds for region 2 (noun) for the French experiment with 95 % confidence intervals. Note that higher reciprocal reading times correspond to lower reading times prior to transformation.

$SE = 0.00006, t = -4.62, p < 0.001$), showing that the reading times of 3rd graders are significantly higher. Figure 5 displays the transformed data. Figures of the untransformed data for regions 2 and 3 as well as the mean reading times and standard deviations per level can be found in the Supplementary material.

4.5.3 Region of interest 3 – spillover region

A linear mixed effects model was fit to the reciprocal reading times on the preposition, which functioned as the spillover region. The model included *Grammaticality*, *Education*, and *Congruency* as main effects, as well as two-way interaction effects between

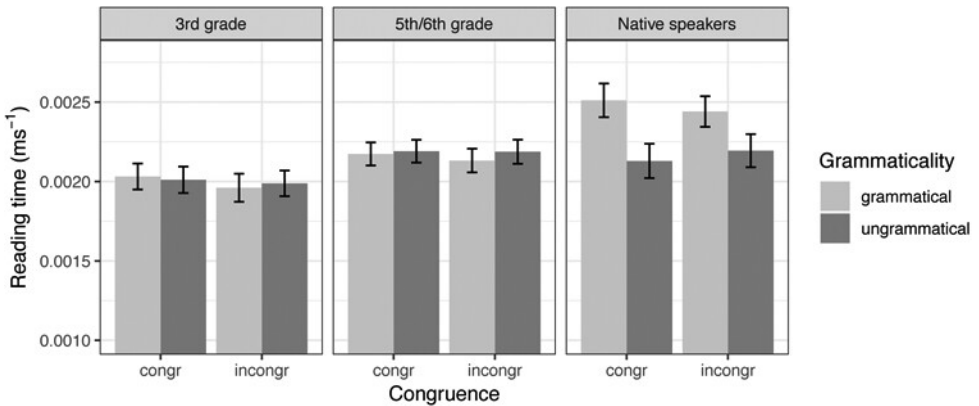


Figure 6 Reading times in reciprocal milliseconds for region 3 (preposition) for the French experiment with 95% confidence intervals. Note that higher reciprocal reading times correspond to lower reading times prior to transformation.

Grammaticality and *Education* on the one hand, and *Congruency* and *Education* on the other hand. Random intercepts were estimated per participant and per item. In addition, random slopes for the first *Education* contrast (3rd graders vs. average) were included per item, as well as the interaction term between the same *education* contrast and *Grammaticality*. There was a significant effect of *Education* when comparing the 3rd graders to the average ($\beta = -0.00017$, $SE = 0.00004$, $t = -3.741$, $p < 0.001$), with 3rd graders' reading times being significantly higher. There was also a significant main effect of *Grammaticality*, with ungrammaticality leading to higher reading times ($\beta = 0.00005$, $SE = 0.00001$, $t = 4.08$, $p < 0.001$). Finally, we found a significant interaction effect involving *Grammaticality* and *Education* when comparing the 3rd graders to the average ($\beta = -0.00004$, $SE = 0.00002$, $t = -2.61$, $p = 0.009$) and when comparing the 5th/6th graders to the average ($\beta = -0.00006$, $SE = 0.00002$, $t = -4.13$, $p < 0.001$). The transformed data are displayed in Figure 6, demonstrating that *Grammaticality* only affected the reading times of the L1 speakers of French.

4.6 Discussion

The results indicate that the grammaticality of the sentences affected reading times, but only among the native speakers of French. Recall that we hypothesized that Dutch learners of French would experience processing difficulty when reading French sentences with a mismatch between the semantic gender of the possessor noun and the syntactic gender of the possessive pronoun (e.g. *Caroline a acheté son piano à Londres* 'Caroline bought her.M piano.M in London'), because they experience such difficulties in their own language. However, no systematic differences in reading times were observed among the Dutch learners of French, and thus the effect we found in Experiment 1 is not

replicated for L2 processing. As noted, Dutch learners of French were also not sensitive to ungrammaticality in their L2. The only effect that reached significance regarding the Dutch learners is that of *Education*, indicating that Dutch learners of French read more slowly than native speakers of French, but this finding is not surprising nor relevant for our hypotheses.

Regarding the absence of the expected effects, we may wonder whether the Dutch learners understood the French sentences they read at all, as they were not very accurate at answering the control statements. In fact, we had to remove half of the control statements because they were answered incorrectly by the majority of Dutch participants. Yet, the difference in reading times between participant groups suggests that they did in fact pay attention to what they were reading, and the participants were moreover able to provide the correct gender of the nouns in the decision task that followed. The low accuracy in the control statements is perhaps due to the statements themselves. The (ungrammatical) sentence *Félix a cherché son chaussure dans l'armoire* 'Felix looked for his shoe in the closet', for example, was followed by the control statement *Félix a perdu une chaussure* 'Felix lost a shoe' with the intended answer 'correct'. However, the control statement contains the verb *perdu* 'lost', while in Dutch one would use an adjective *kwijt* 'lost' here, focusing on the resulting state of not having the shoe rather than on the activity of losing it. Dutch participants possibly interpreted this French statement in such a way that Felix 'actively' lost his shoe, which is not necessarily the case if he simply cannot find it in the closet. Participants may have consequently considered this statement incorrect. Thus, while the set of control statements was clearly not optimal to test whether our Dutch high school students understood the French sentences, we still have sufficient reason to assume they did.

The absence of gender congruency effects among the learners of French may have been expected based on Versendaal (2013), as she did not find such an effect in her experiment either and links the absence of such effects to *good enough processing* (Ferreira et al., 2002). But the insensitivity of Dutch learners of French to ungrammaticality in their L2 was not necessarily expected. Previous studies did find processing effects of morphosyntactic gender violations by L2-speakers, but not when these learners were low proficient (e.g., Keating 2009; Sagarra & Herschensohn, 2010). Sagarra and Herschensohn (2010) tested the effects of syntactic gender violations in beginning and intermediate adult English-speaking learners of Spanish as well as in Spanish native speakers. They found that beginning learners were completely insensitive to these violations, whereas intermediate learners did show sensitivity and had significantly longer reading times for syntactically disagreeing adjectives than for agreeing ones.

The intermediate L2-learners in Sagarra and Herschensohn (2010) were seventh- or eighth-semester students of Spanish at a North-American university. Their L2 proficiency level was therefore much higher than that of the high school students in our experiment. Versendaal (2013) found a sharp decrease in reading speed between Dutch

sixth year high school students and first year university students of French, as well as between Dutch first year university students and native speakers of French. Indeed, despite having had some years of learning French, the proficiency level of fifth and sixth year high school students is still quite low, and in any case much lower than that of first year (i.e. first or second semester) university students of French in the Netherlands (cf. Versendaal 2013). The learners of French in our experiment must therefore be considered beginning learners, which may explain why we did not find any grammaticality effects in their reading times (cf. Sagarra & Herschensohn, 2010). We leave the question whether university students are sensitive to this contrast to future research.

5 General discussion

Dutch natives show an effect of congruency between the gender of the subject and a subsequent possessive pronoun in their L1 (Experiment 1). Structures with such a gender mismatch lead to a slowdown in reading times. This finding is in line with a previous judgment experiment (Piepers & Redl 2018) and with the experimental literature (Carreiras et al., 1996; Dong et al., 2015; Nieuwland, 2014; Lago et al., 2019).

However, the congruency effect found in Experiment 1 is not reflected in the data from Experiment 2, in which we tested the effect of congruency between the semantic gender of the subject and the syntactic gender of the possessive pronoun in French, syntactically (dis)agreeing with the possessee. Our hypothesis was that Dutch learners of French might process the syntactic gender feature as a semantic one. This is not what we found. Unlike Lago et al. (2019), we did not find an effect of an infelicitous gender combination of the subject and the possessive pronoun. The crucial difference between the experiment conducted by Lago et al. (2019) and ours is that their infelicitous sentences in German were indeed infelicitous, like the infelicitous Dutch sentences in our Experiment 1, but the French sentences in our Experiment 2 were sometimes ungrammatical, yet never infelicitous, because the possessive pronoun in French does not have a semantic gender distinction between 'his' and 'her'.

We did not find an effect of ungrammaticality among the Dutch learners either, which is presumably related to their low level of French. The Dutch learners of French in our experiment were high school students who are approximately 14 to 18 years old and only study French for school purposes. As suggested in Versendaal (2013), the absence of processing difficulties due to syntactic grammaticality violations among Dutch learners of French can be explained in terms of *good-enough*, *underspecified*, or *shallow language processing* (Christianson, 2016; Christianson et al., 2001; Ferreira et al., 2002; Ferreira & Lowder, 2016; Ferreira & Patson, 2007; Sanford & Sturt, 2002). L2 learners have been found to underutilize syntactic information during sentence processing in comparison to native speakers (Keating, 2009). Although L2 learners can achieve native-like processing

of syntactic gender agreement violations, as tested in our second experiment, this has only been found with highly proficient advanced or intermediate learners (Keating, 2009; Sagarra & Herschensohn, 2010; Clahsen & Felser, 2018).

6 Conclusion

In a self-paced reading experiment with native speakers of Dutch from secondary education (14–18 years old) we measured an increase in reading times in case of a semantic gender mismatch between the subject of a sentence and a subsequent possessive pronoun in Dutch. We hypothesized that a similar effect would be present for Dutch high school students of French in case of a perceived mismatch between the semantic gender of the subject and the syntactic (unrelated) gender of the possessive pronoun in French. However, we did not find such an effect nor did we find an effect of ungrammaticality, which can be explained by the beginning learners' level of French, leading to shallow language processing.

Acknowledgements

We would like to thank Marjolein Klaassen for her help collecting the data and Marine Vuillermet for her help finding native French participants. We are also grateful to the staff and students of the schools that participated in the experiments: Stedelijk College Eindhoven, Jacob-Roelandslyceum in Boxtel, and Carmelcollege in Emmen.

Notes

- 1 The masculine pronoun *son* is also used when the following word starts with a vowel or a *mute h*, which may also lead to a gender mismatch. We will not discuss these cases in this paper.

References

- Antón-Méndez, I. (2011). Whose? L2-English speakers' possessive pronoun gender errors. *Bilingualism: Language and Cognition*, 14(3), 318–331. <https://doi.org/10.1017/S1366728910000325>
- Baayen, R.H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28. <https://doi.org/10.21500/20112084.807>
- 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(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>

- Bates, D., Kliegl, R., Vasishth, S., & Baayen, R.H. (2015a). Parsimonious mixed models. *ArXiv Preprint ArXiv:1506.04967*, 1–27. <https://arxiv.org/abs/1506.04967>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *The Quarterly Journal of Experimental Psychology*, 69(5), 817–828. <https://dx.doi.org/10.1080/17470218.2015.1134603>
- Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psycholinguistics*, 49(3), 639–663. <https://doi.org/10.1080/173755647>
- Christianson, K., Hollingworth, A., Halliwell, J.F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42(4), 368–407. <https://doi.org/10.1006/cogp.2001.0752>
- Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, 40(3), 693–706. <https://doi.org/10.1017/S027226317000250>
- Corbett, G.G. (1979). The agreement hierarchy. *Journal of Linguistics*, 15(2), 203–224. <https://doi.org/10.1017/S0022226700016352>
- Dong, Y., Wen, Y., Zeng, X., & Ji, Y. (2015). Exploring the cause of English pronoun gender errors by Chinese learners of English: Evidence from the self-paced reading paradigm. *Journal of Psycholinguistic Research*, 44(6), 733–747. <https://doi.org/10.1007/s10936-014-9314-6>
- Ferreira, F., Bailey, K.G.D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15. <https://doi.org/10.1111/2F1467-8721.00158>
- Ferreira, F., & Lowder, M.W. (2016). Prediction, information structure, and good-enough language processing. *Psychology of Learning and Motivation*, 65, 217–247. <https://doi.org/10.1016/bs.plm.2016.04.002>
- Ferreira, F., & Patson, N.D. (2007) The ‘Good Enough’ approach to language comprehension. *Language and Linguistics Compass*, 1(1–2), 71–83. <https://doi.org/10.1111/j.1749-818X.2007.00007.x>
- Keating, G.D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, 59(3), 503–535. <https://doi.org/10.1111/j.1467-9922.2009.00516.x>
- Kliegl, R., Masson, M.E.J., & Richter, E.M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, 18(5), 655–681. <http://doi.org/10.1080/13506280902986058>
- Lago, S., Stutter Garcia, A., & Felser, C. (2019). The role of native and non-native grammars in the comprehension of possessive pronouns. *Second Language Research*, 35(3), 319–349. <https://doi.org/10.1177/0267658318770491>
- Mitchell, D.C. (1984). An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In D.E. Kieras, & M.A. Just (Eds.), *New methods in reading comprehension research* (pp. 69–89). Erlbaum.
- Nieuwland, M.S. (2014). “Who’s he?” Event-related brain potentials and unbound pronouns. *Journal of Memory and Language*, 76, 1–28. <https://doi.org/10.1016/j.jml.2014.06.002>

- Piepers, J., & Redl, T. (2018). Gender-mismatching pronouns in context. In B. LeBruyn, & J. Berns (Eds.), *Linguistics in the Netherlands* 35 (pp. 97–110). John Benjamins. <https://doi.org/10.1075/avt.00007.pie>
- Pozzan, L., & Antón-Méndez, I. (2017). English possessive gender agreement in production and comprehension: Similarities and differences between young monolingual English learners and adult Mandarin–English second language learners. *Applied Psycholinguistics*, 38(4), 985–1017. <https://doi.org/10.1017/S0142716417000017>
- R Core Team. (2018) *R: A language and environment for statistical computing*. R foundation for statistical computing. <https://www.r-project.org/>
- Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, 120(8), 2022–2039. <https://doi.org/10.1016/j.lingua.2010.02.004>
- Sanford, A.J., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Science*, 6(9), 382–386. [https://doi.org/10.1016/s1364-6613\(02\)01958-7](https://doi.org/10.1016/s1364-6613(02)01958-7)
- Slevc, L.R., Wardlow Lane, L., & Ferreira, V.S. (2007). Pronoun production: Word or world knowledge? *MIT Working Papers in Linguistics*, 53, 191–203.
- Versendaal, R. (2013) Onderzoeksrapport Honours-onderzoek: Een onderzoek naar de verwerking van het bezittelijk voornaamwoord door Nederlandse leerders van het Frans [Research report Honours-study: An investigation of the processing of the possessive pronoun by Dutch learners of French]. Unpublished manuscript, Radboud University Nijmegen.
- White, J., Muñoz, C., & Collins, L. (2007). The *his/her* challenge: Making progress in a ‘regular’ L2 programme. *Language Awareness*, 16(4), 278–299. <https://doi.org/10.2167/la419.0>