The role of native and non-native grammars in the comprehension of possessive pronouns

Sol Lago, Anna Stutter Garcia & Claudia Felser
Potsdam Research Institute for Multilingualism, Potsdam, Germany

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

Abstract

Previous studies have shown that multilingual speakers are influenced by their native (L1) and non-native (L2) grammars when learning a new language. But so far, these studies have mostly used untimed metalinguistic tasks. Here we examine whether multilinguals' prior grammars also affect their sensitivity to morphosyntactic constraints during processing. We use speeded judgment and self-paced reading tasks to examine the comprehension of German possessive pronouns. To investigate whether native and non-native grammars differentially affect participants' performance, we compare two groups of non-native German speakers with inverse L1-L2 distributions: a group with L1 SPANISH-L2 ENGLISH, and a group with L1 ENGLISH-L2 SPANISH. We show that the reading profiles of both groups are modulated by their L1 grammar, with L2 proficiency selectively affecting participants' judgment accuracy but not their reading times. We propose that reading comprehension is mainly influenced by multilinguals' native grammar, but that knowledge of an L2 grammar can further increase sensitivity to morphosyntactic violations in an additional language.

Keywords

I Introduction

When speakers learn an additional language (La), they have available at least two sources of prior knowledge: their native grammar (L1) and the grammar of any non-native languages (L2s) acquired after their mother tongue. While much research has examined the role of L1 and L2 grammars in La production (Bohnacker, 2006; Bardel & Falk, 2007; Flynn, Foley & Vinnitskaya, 2004; García Mayo, Ibarrola & Liceras, 2005; Leung, 2003, 2005, 2007), comprehension studies have only recently been conducted (Amaro, Amaro & Rothman, 2015; Falk & Bardel, 2011; Hermas, 2010, 2014; Rothman & Cabrelli Amaro, 2010; Sanz, Park & Lado, 2015; Westergaard, Mitrofanova, Mykhaylyk & Rodina, 2017). But so far, comprehension work has mostly used untimed metalinguistic paradigms, and little is known about how prior grammatical knowledge affects La online processing.

The current study seeks to bridge this gap by examining whether multilingual speakers' prior grammars affect their comprehension of pronouns in real time. We focus on possessive pronoun forms[1] such as *his* and *her*, which are notoriously problematic for L2 learners (Zobl, 1985; Martens, 1988; Lightbown & Spada, 1990; Spada, Lightbown & White, 2005; Muñoz, 1994; White, Muñoz, & Collins, 2007). For instance, French and Spanish students of English often produce possessives with incorrect gender marking, such as *The dad put <u>her</u> little girl on his shoulder* (White, 1998). Here we investigate whether such errors are mirrored in judgment and comprehension tasks, and whether they extend to proficient multilingual speakers. Since the morphosyntactic constraints on possessive pronouns vary cross-linguistically, they provide a good test case for examining the influence of native vs. non-native grammars in multilingual processing.

II Possessive gender agreement

Possessives need to establish referential and morphosyntactic relationships with other elements in a sentence. Here we focus on gender agreement, which varies cross-linguistically. For instance, the German possessives *sein/ihr* ('his/her') simultaneously agree in gender with a preceding possessor and a following possessee noun. Possessor agreement is carried by the stem of the pronoun (*sein* for masculine possessors vs. *ihr* for feminine possessors) whereas possessee agreement is carried by the suffix (*–en* for masculine accusative vs. *–e* for feminine forms):

(1)   a. **Frau Müller**.possessor liebt <u>ihren</u> **Sohn**.   German
          Ms. Müller loves her.masc *son.*

       b. **Herr Müller**.possessor liebt <u>seine</u> **Tochter**.
          Mr. Müller loves his.fem *daughter.*

By contrast, English possessives only agree in gender with the possessor noun (2), whereas in Spanish and many other Romance languages, gender agreement depends solely on the possessee noun (3):

---

[1] We use the term "possessive pronoun" to emphasize that forms like *his/her* require an antecedent to receive an interpretation. Syntactically, possessive pronouns function as modifiers of a head noun and inherit its agreement features, similarly to determiners and adjectives.

(2) a. **Ms. Müller**.possessor loves <u>her</u> son. English

    b. **Mr. Müller**.possessor loves <u>his</u> daughter.

(3) a. Nosotros queremos a <u>nuestra</u> **hija**.possessee. Spanish
      We.masc *love our*.fem *daughter.*

    b. Nosotras queremos a <u>nuestro</u> **hijo**.possessee.
      We.fem *love our*.masc *son*

Note that in contrast with (3a,b), the third person Spanish possessive *su* does not explicitly mark gender agreement. This absence likely reflects a morphological loss rather than a lack of underlying syntactic agreement, because possessee agreement is still overtly realized in the first and second grammatical persons (as shown in 3a,b) and it occurs with number features (e.g. *su*.sg *hijo* vs. *sus*.pl *hijos*). Thus, the examples above show that possessive gender agreement in Spanish is forward-looking because it involves a relationship with a following possessee noun. In contrast, English possessive agreement is backward-looking because it involves a relationship with a noun-phrase external possessor noun.

These cross-linguistic differences may create problems for language learners, such that the use of possessives might be challenging when agreement constraints differ between their native and non-native languages. This possibility has previously been examined in second language speakers (Antón-Méndez, 2011; Pozzan & Antón-Mendez, 2017; Santesteban, Foucart, Pickering & Branigan, 2010; Foucart, Santesteban, Branigan & Pickering, 2011). For instance, Antón-Méndez conducted a production study with Spanish, Italian and Dutch proficient speakers of English. These languages differ in their agreement constraints: Dutch and English require possessor agreement, whereas Spanish and Italian require possessee agreement.

Antón-Méndez hypothesized that if participants' native agreement constraints affected their production of L2 possessives, Spanish and Italian (but not Dutch) speakers should make possessive gender errors in English when the possessor and possessee noun mismatched in gender (e.g. *John _____ sister,* henceforth a MATCH effect). This prediction was borne out: Romance speakers made more errors than Dutch speakers when the possessor and possessee mismatched in gender, by producing possessives that matched the gender of the possessee noun. Further, Spanish and Italian speakers erred at similar rates, suggesting that the lack of overt gender marking in Spanish third person possessives did not prevent L1 Spanish speakers from erring in English. Thus, Antón-Méndez proposed that Romance speakers had problems inhibiting their native forward-looking agreement mechanism during L2 production.

A subsequent study found the same type of gender errors in the production of 3–5-year-old English-speaking children and adult native speakers of Chinese (Pozzan & Antón-Méndez, 2017). Since these learners' errors could not be explained by L1 influence (as both languages lack possessee agreement), the authors suggested that erroneously computing agreement with a following possessee reflected an universal tendency to establish agreement locally within the noun phrase.

Pozzan and Antón-Méndez also conducted a comprehension act-out task, where learners heard sentences such as *Give the apple to his sister* and had to perform the action in a visual display. The display showed two adult characters (a male and a female protagonist), each with a male and female sibling. Participants' eye movements were recorded to examine whether they predictively directed their gaze to a gender-matching possessee upon hearing the gender-marked possessive (e.g. a male sibling in the example above), as would be expected if they wrongly computed gender in a forward-looking manner. However, Chinese speakers made almost no act-out errors, and the (mis)match between the possessor and possessee nouns did not affect their gaze patterns.

These results suggest that gender agreement errors with English possessives occur in production but not comprehension, and that production errors may arise from a universal tendency towards local agreement computations rather than from L1 influence. Alternatively, the lack of comprehension errors might have resulted from the properties of the act-out task. The task, which was designed to be performed by children, may have been too easy for the adult L2 participants, who had unlimited time to act out the sentences. Further, the comprehension task was always performed after the production task, providing participants with prior practice with possessives. Finally, Chinese learners were not compared to another L2 group whose L1 had possessee agreement. Therefore, these results show that possessive errors cannot be solely due to L1 influence, but they do not address the question of whether errors are more pervasive in speakers of languages with varying agreement constraints. The current study revisits this question by using two comprehension tasks that put participants under processing pressure in order to increase the likelihood of errors. We compare multilingual speakers whose prior grammars either have or lack possessee agreement, to address whether possessive errors result from L1 transfer or from a generalized tendency towards local agreement.

III   Models of La acquisition

Existing accounts of La acquisition focus on beginning language learners and on whether their hypotheses about the La grammar are more affected by their L1 or L2 knowledge (for review, see González Alonso & Rothman, 2017; González Alonso, Rothman, Berndt, Castro & Westergaard, 2016; Rothman & Halloran, 2013; de Bot & Jaensch, 2015; Jaensch, 2013). Several studies have looked at whether a native language, which has been acquired earlier and used for longer, is more internalized than a non-native language and thus more likely to be used to generate hypotheses about La grammatical constraints (Hermas, 2010, 2014; Lozano, 2002; Jin, 2009; Na Ranong & Leung, 2009). These studies have found that learners' L1 affects La acquisition, such that La morphosyntactic constraints are more likely to be successfully acquired when already instantiated in the L1. For example, learners whose L1 licenses null arguments are more likely to apply the null subject constraint in an La than learners whose L1 lacks null subjects (Lozano, 2002).

By contrast, other accounts propose that speakers' L2 knowledge affects La acquisition to a larger extent than their L1 (Bardel & Falk, 2007, 2012; Falk & Bardel, 2011; Falk, Lindqvist, & Bardel, 2015). These accounts observe that L2 grammars are acquired in a metalinguistic manner, frequently with some explicit language instruction or in classroom settings. As a result, learners might store non-native grammars in a different memory system than the native grammar (e.g. Ullman, 2001). Due to their non-native

status, learners might perceive the L2 and La grammars as more similar, making the L2 grammar more available for learning a La.

Finally, a third type of account proposes that learners transfer a language depending on its typological similarity with the La, regardless of its native or non-native status. These accounts are divided between those that posit that similarity is established on a grammatical property-by-property basis (Flynn et al., 2004; Slabakova, 2017; Westergaard et al., 2017) and those that argue that learners perceive similarity more globally based on the lexical, phonological, and morphosyntactic parallels between the prior languages and the La (Rothman, 2010, 2013, 2015). For instance, learners of German who know Spanish and English might perceive English and German as typologically closer due to the strong similarity between their phonological and lexical systems: German shares more vocalic and consonantal phonemes with English than Spanish, and both languages have freer stress placement and a preference for closed syllable structures, with longer average durations and a higher consonant-to-vowel ratio per syllable than Spanish (Delattre, 1965, 1969). At the word level, German and English are also more similar than German and Spanish, with the former sharing more translation equivalents as well as phonetic and orthographic cognates (Schepens, Dijkstra, & Grootjen, 2012; Schepens, Dijkstra, Grootjen, & van Heuven, 2013). Therefore, due to the stronger similarity between English and German (as compared to Spanish and German), similarity accounts predict that speakers should transfer their English grammar to the learning of German, regardless of its native or non-native status. In contrast, property-by-property similarity accounts predict that learners' acquisition will depend on the similarity of the target La construction. As both accounts make similar predictions for this study, they are discussed jointly under the term "similarity-based models".

As mentioned previously, existing La models deal with acquisition rather than processing and mostly focus on the early stages of La learning and on learners' initial hypotheses about the nature of La syntactic constraints. The processing of problematic constructions may however continue to be challenging even for learners at higher proficiency levels. We address this possibility by examining whether proficient multilinguals show processing difficulties with constructions that they have successfully acquired, and whether their difficulties are differentially modulated by their L1 and L2 grammars.

IV   The current study

We examined the processing of possessor agreement with the German pronouns *sein* and *ihr*. With this goal, we created felicitous and infelicitous sentence pairs by manipulating gender agreement between the pronoun and the possessor noun phrase, as shown in (4a,b):

(4)   a. Frau Schmidt küsste ihre Mutter bei dem letzten Familientreffen.
           Ms. Schmidt kissed her mother at the last family reunion.

        b. #Frau Schmidt küsste seine Mutter bei dem letzten Familientreffen.
           Ms. Schmidt kissed his mother at the last family reunion.

Note that (4b) is not strictly ungrammatical: it is grammatical under the reading that the mother was not Ms. Schmidt's but that of a male person outside the current sentence

context. But (4b) is clearly infelicitous under the reading that the mother was Ms. Schmidt's. Previous research has shown that these sentences elicit processing disruptions, because comprehenders initially try to establish coreference with sentence-internal antecedents (Chow, Lewis, & Phillips, 2014; Niewland, 2014; Lamers, Jansma, Hammer, & Münte, 2006; Osterhout & Mobley, 1995). When these antecedents are inaccessible, comprehenders show processing difficulty, either because the sentences are deemed anomalous or because postulating a sentence-external antecedent takes additional time and processing effort.

We investigated whether multilingual speakers differed in their sensitivity to possessor agreement violations depending on whether possessor agreement was present in their L1. We selected non-native speakers of German who knew both Spanish and English. One group consisted of Spanish native speakers who had learned English as a non-native language (L1 SPANISH-L2 ENGLISH). The other group consisted of English native speakers who had learned Spanish as a non-native language (L1 ENGLISH-L2 SPANISH). In what follows, we refer to the native language as L1, the non-native language as L2 and the non-native language under study, German, as La. Note that the L2-La terms do not necessarily indicate temporal order (as it does e.g. in Jaensch, 2013). Rather, La is used for "additional language" or LX (Van der Silk, van Hout & Schepens, 2017; García Mayo & Rothman, 2012; Dewaele & Salomidou, 2017). We adopt this terminology because our study was not designed to address the role of temporal order in acquisition, but rather, whether a native and a non-native grammar differentially affected the processing of an additional non-native language.

By testing these two groups of La German speakers, we investigated whether Spanish native speakers, whose L1 lacked possessor gender agreement, performed worse than English native speakers with possessor agreement in German. Additionally, we examined the potential effects of L2 influence and typological similarity. We hypothesized that if participants were influenced by their L2 agreement constraints, then their L2 proficiency should further modulate their performance or even cause a reversal, with Spanish speakers outperforming English speakers. Finally, if participants were influenced by the typologically closest grammar, there might be no between-group differences, as both La groups should be able to deploy English agreement constraints, regardless of whether these were present in their L1 or L2. Note that the lack of between-group differences would also be consistent with no transfer from any language, as predicted by accounts that attribute possessive errors to a universal/L1-independent tendency to compute agreement locally (Pozzan & Antón-Mendez, 2017). Participants' processing of possessor agreement was tested first with a metalinguistic task (Experiment 1) and then with an implicit task (Experiment 2).

V Experiment 1

Experiment 1 consisted of a speeded acceptability judgment task. Speeded acceptability judgments provide a time-sensitive measure that has been shown to reliably mirror processing effects by requiring participants to rely on their working memory to construct a representation of the sentence and by restricting the amount of time that they have to reflect on their acceptability intuitions (Drenhaus, Saddy, & Frisch, 2005; Wagers, Lau, & Phillips, 2009; Parker & Phillips, 2016).

We tested sentences such as (5a-d), using a 2 (*felicitous/infelicitous*) × 2 (*match/mismatch*) design. In the felicitous conditions, the pronoun and possessor agreed in gender, whereas in the infelicitous conditions they disagreed: (5a) and (5b) vs. (5c) and (5d). In the gender-match conditions, the possessor and possessee had the same gender, whereas in the gender-mismatch conditions they did not :(5a) and (5c) vs. (5b) and (5d).

(5) a. *Felicitous match*

Frau Schmidt küsste <u>ihre</u> Mutter bei dem letzten Familientreffen.
Ms. Schmidt kissed her mother at the last family reunion.

   b. *Felicitous mismatch*

Herr Schmidt küsste <u>seine</u> Mutter bei dem letzten Familientreffen.
Mr. Schmidt ... his mother...

   c. *Infelicitous match*

#Frau Schmidt küsste <u>seine</u> Mutter bei dem letzten Familientreffen.
Ms. Schmidt ... his mother…

   d. *Infelicitous mismatch*

#Herr Schmidt küsste <u>ihre</u> Mutter bei dem letzten Familientreffen.
Mr. Schmidt ... her mother…

We measured acceptability judgments and response times of correctly answered trials. With regard to judgments, we assessed whether English and Spanish speakers were differentially able to detect infelicitous sentences. Further, based on previous production studies, we examined whether there were more judgment errors in mismatch configurations, where the gender of the possessor and possessee noun disagreed, thus increasing Spanish participants' likelihood of L1 influence. If so, we expected Spanish speakers to wrongly accept infelicitous sentences (and to wrongly reject felicitous ones) more often in mismatch configurations. With regard to response times, we predicted that both groups should judge infelicitous sentences more quickly than felicitous sentences, because infelicitous sentences contained a mid-sentence gender-mismatching pronoun that provided a clear and early cue to guide participants' judgments. Further, Spanish natives might take longer to reject infelicitous sentences as compared with English natives, whose processing routines for establishing possessor gender agreement may be more firmly in place due to their L1.

1 Participants

Participants consisted of Spanish and English native speakers with La German.[2] The L1 Spanish group consisted of thirty-one Spanish natives with L2 English and La German (mean age = 29 years, age range = 21–44, 24 females, 27 right-handed). With regard to the order of acquisition of the non-native languages, the majority of Spanish natives had acquired English before German (n = 27). The L1 English group consisted of thirty-four

---

[2] In order to assess the validity of the materials and experimental design, the experiments reported were also performed by a group of native German speakers. In all experiments German speakers behaved as predicted and showed clear effects of felicitousness (see *Appendix I*).

English natives with L2 Spanish and La German (mean age = 27 years, age range = 19–56, 21 females, 32 right-handed). Sixteen English natives had acquired Spanish before German, whereas the remaining eighteen participants had acquired them simultaneously or in the inverse order. Table 1 summarizes the demographic profile of the La groups, who were matched in their German age of acquisition and proficiency. Participants rated their L2 and German proficiency in each of the four skills (speaking, listening, reading and writing) using a scale from 1–10. These ratings were then averaged to get an overall measure of their language proficiency. Self-ratings were used because previous research has shown that they offer a good correspondence with formal language tests (Blanche & Merino, 1989; Ross, 1998; Marian, Blumenfeld, & Kaushanskaya, 2007), they can be obtained quickly, and they avoid the problem of comparing tests that may vary in the design and level of difficulty of their materials across languages.

In addition to self-ratings, participants' knowledge of German was gauged with the Goethe Institute Placement Test (Goethe Institute, 2010). Participants' Goethe scores were highly consistent with their German self-ratings, with both groups showing significant correlations between these measures (*L1 Spanish*: $r = .52$, $t = 3.280$, $p = .003$; *L1 English*: $r = .49$, $t = 3.196$, $p = .003$). To ensure that participants' overall knowledge of German corresponded to an intermediate-to-advanced level, only speakers who scored 50% or higher in the Goethe test were included in the analyses.

Participants had acquired their L2 non-natively and had at most basic knowledge of other foreign languages. In this and following experiments, participants provided informed consent and received financial compensation for their participation. All procedures were in accordance with the Declaration of Helsinki.

Table 1. Demographic profiles of the multilingual participants in Experiments 1 and 2 (standard deviations in parentheses).

	Experiment 1		Experiment 2	
	L1 Spanish	*L1 English*	*L1 Spanish*	*L1 English*
German age of acquisition (years)	19 (7)	17 (6)	21 (8)	20 (6)
German self-rated proficiency (%)	77 (12)	74 (16)	73 (13)	75 (15)
Listening	80 (14)	79 (18)	79 (14)	80 (15)
Speaking	74 (15)	71 (17)	70 (16)	70 (18)
Reading	80 (14)	78 (17)	79 (13)	80 (16)
Writing	73 (13)	69 (20)	64 (18)	68 (18)
German Goethe score (%)	73 (9)	75 (11)	75 (13)	72 (13)
L2 age of acquisition (years)	8 (3)	15 (5)	8 (3)	14 (6)
L2 self-rated proficiency (%)	81 (11)	59 (19)	81 (11)	58 (22)
Listening	81 (14)	58 (21)	84 (11)	63 (21)
Speaking	77 (13)	56 (21)	76 (15)	50 (23)
Reading	86 (11)	67 (22)	87 (10)	66 (25)
Writing	77 (14)	56 (22)	76 (15)	52 (26)

2 Materials

Materials consisted of 32 experimental items, 30 fillers and 32 items from a separate experiment (not reported here). Each item contained a possessive pronoun and an animate possessor noun. The gender of the possessor was indicated by the preceding title *Frau/Herr* ('Ms./Mr.') to avoid gender ambiguity. The pronoun was always followed by an animate possessee noun, whose gender was marked lexically (e.g. "*Mutter*", 'mother') or morphologically (e.g. "*Lehrer-in*', 'teacher-fem'). In half of the items the possessee was female, and in half it was male. The pronoun always showed morphosyntactic gender agreement with the possessee noun (as indicated by its suffix). Across the experiment, half of the infelicitous sentences contained a feminine pronoun and half contained a masculine pronoun, to prevent participants from using the pronoun's gender as a cue to the felicitousness of the sentence. All experimental materials, as well as data for this and following experiments, are available at the Center for Open Science Framework website (https://osf.io/).

Half of the fillers were grammatical and half were ungrammatical. The fillers were constructed to ensure that differences between the experimental conditions did not result from overall differences between the groups. Thus, the fillers tested phenomena that were relevant to both the grammars of Spanish and English: word order (n = 10), subject-verb number agreement (n = 10) and tense correlations (n = 10). Across an experimental session, participants saw ungrammatical (or infelicitous) and grammatical (or felicitous) sentences in a 1:1 ratio. The experimental items were distributed across four lists in a Latin-square design. Experimental and filler items were randomized on a by-participant basis.

3 Procedure

Participants were asked to judge whether the sentences sounded acceptable in German. Sentences were presented one word at a time in the center of the screen with a presentation rate of 500 ms per word. After each sentence, participants saw the question *Was the sentence acceptable?* and had to provide a response using either the F ("acceptable") or J key ("unacceptable") within a timeout of 3000 ms. The task was run on the web-based platform Ibex Farm (Drummond, 2013). Web-based testing was used because it allowed us to expand our participant pool by recruiting La speakers worldwide, and because this method has been found to yield reliable results in previous psycholinguistics studies (Enochson & Culbertson, 2014; Sprouse, 2011; Gibson, Piantadosi, & Fedorenko, 2011; Dillon, Clifton, & Frazier, 2014; Chemla, Cummins & Singh, 2015; Wagers & Phillips, 2014).

Before beginning the experimental session, participants filled in a demographic survey. Then they were given instructions and completed three practice trials with feedback. The instructions defined the terms 'acceptable' and 'unacceptable' (i.e. a sentence that a German speaker would/would not say) and provided participants with examples. To discourage participants from considering unmentioned potential referents, they were instructed to provide their judgments only based on the current sentence context.

After the speeded acceptability task, La participants completed the Goethe Institute Placement Test, a vocabulary test and an untimed choice task. The goal of the vocabulary test was to ensure that participants understood the words used in the speeded task, including the possessee noun (e.g. *mother* in (5)) and the verb (e.g. *kissed*). The goal of the untimed choice task was to ensure that participants knew possessor agreement constraints in

German, so that any effects in the speeded task were not attributable to incomplete grammatical knowledge. The post-test consisted of 60 items, of which 10 tested for participants' knowledge of possessor agreement. These items were similar to those in the online task. Half of the items were felicitous (e.g. *Herr Neumann bezahlte seine Köchin*, 'Herr Neumann paid his cook') and half infelicitous (e.g. *#Herr Hoffmann suchte ihre Nachbarin*, 'Herr Hoffmann looked for her neighbor'). Following previous work, a threshold of 75% accuracy or higher was used to determine mature knowledge (Foote, 2009; Hsien-jen Chin, 2009; Falk & Bardel, 2011). An experimental session lasted approximately 70 minutes.

4   Analysis

We analyzed the proportion of sentences judged as acceptable and the response times of correctly answered trials. Acceptability data were analyzed with mixed-effects logistic regression (Jaeger, 2008) and response times were analyzed with mixed-effects linear models (Baayen, Davidson & Bates, 2008). The logistic transformation was used for the acceptability data. Following the Box-Cox procedure (Box & Cox, 1964), response times were reciprocally transformed. Items with unknown vocabulary were excluded on a by-participant basis, resulting in the exclusion of 5.36% of the data for the L1 Spanish group and 3.88% of the data for the L1 English group. To ensure that the analysis only included participants who were able to perform proficiently in the speeded task, participants with accuracy lower than 60% in the filler trials were excluded ($n_{L1\ Spanish} = 5$; $n_{L1\ English} = 4$). In addition, responses shorter than 200 ms were removed (Staub, 2010), resulting in the exclusion of 4.40% of the data for the L1 Spanish group and 3.64% of the data for the L1 English group. Analyses were performed with R (R Development Core Team, 2017).

The statistical analysis was performed in two stages. The first analysis performed between-group comparisons: acceptability and response time data were analyzed using a model with FELICITOUSNESS (*felicitous/infelicitous*), GROUP (*L1 Spanish/L1 English*) and their interaction as fixed effects. The effect of FELICITOUSNESS assessed whether La speakers were able to detect possessive errors, and the interaction between FELICITOUSNESS and GROUP evaluated whether this ability differed between groups. All categorical factors were coded using sum contrasts.

The second analysis considered each group separately, to examine the role of L2 and German proficiency. We assessed whether responses were modulated by the FELICITOUSNESS of the sentence (*felicitous/infelicitous*), by the gender MATCH between the possessor and possessee noun (*match/mismatch*) and by participants self-rated proficiency in their L2 and in German, which were centered and used as continuous predictors. The order of acquisition of the non-native languages was used as an additional predictor, to address whether having acquired an L2 before or after German affected participants' performance.

For the random effects structure of the models we followed current guidelines in psycholinguistics and initially constructed a maximal model that included random intercepts and slopes for all categorical fixed effects (Barr, Levy, Scheepers & Tily, 2013). When this maximal model failed to converge, the correlation between intercepts and slopes was removed. We report effect sizes using the model estimates ($\hat{\beta}$), standard errors (SE) and *t*- and *z*-statistics. P-values were computed using Satterthwaite's approximation for denominator degrees of freedom (Kuznetsova, Bruun Brockhoff & Haubo Bojesen

Christensen, 2014). Figures and averages are displayed in untransformed measures for easier interpretability, but statistical analyses were always performed on transformed measures, such that differences between conditions may look different in transformed than in untransformed measures.

5 Results

5.1 Group analysis

In the untimed choice task, which tested for knowledge of possessor agreement constraints, participants with lower than 75% accuracy were excluded from analysis. For the remaining 46 participants (21 Spanish and 25 English natives) accuracy did not differ between groups, suggesting that they were similarly aware of possessor constraints when tested offline (*L1 Spanish*: 95%; *L1 English*: 96%, Wilcoxon signed-rank test, W = 279, p = .684). The groups did not differ in their filler accuracy in the online task either (*L1 Spanish*: 80%; *L1 English*: 80%; *ns. effect of Group*: $\hat{\beta}$ = -0.055; *SE* = 0.266; *z* = -0.206; *p* = .837).

Figure 1 shows the acceptability and response time performance of La speakers in the online task. Table 2 shows the results of the statistical analysis. The acceptability data showed an effect of felicitousness, with infelicitous sentences eliciting fewer "acceptable" responses than felicitous sentences. However, the effect did not differ reliably between groups, as shown by the lack of an interaction between felicitousness and group.

The response times of correct trials showed an effect of felicitousness, with faster response times in infelicitous than in felicitous sentences. Crucially, response times also showed an interaction between felicitousness and group, with the effect of felicitousness being stronger for English than for Spanish natives. As shown by Figure 1B, Spanish speakers showed marginally longer response times than English speakers when rejecting infelicitous sentences, suggesting that they were less efficient in using the pronoun's gender to guide their judgments ($\hat{\beta}$ = 0.327; *SE* = 0.181; *t* = 1.808; *p* = .078). Importantly, the increase in response times was not due to an overall slowness of the Spanish group, as their response latencies for felicitous sentences did not differ from English speakers' responses ($\hat{\beta}$ = -0.173; *SE* = 0.184; *t* = -0.940; *p* = .353).
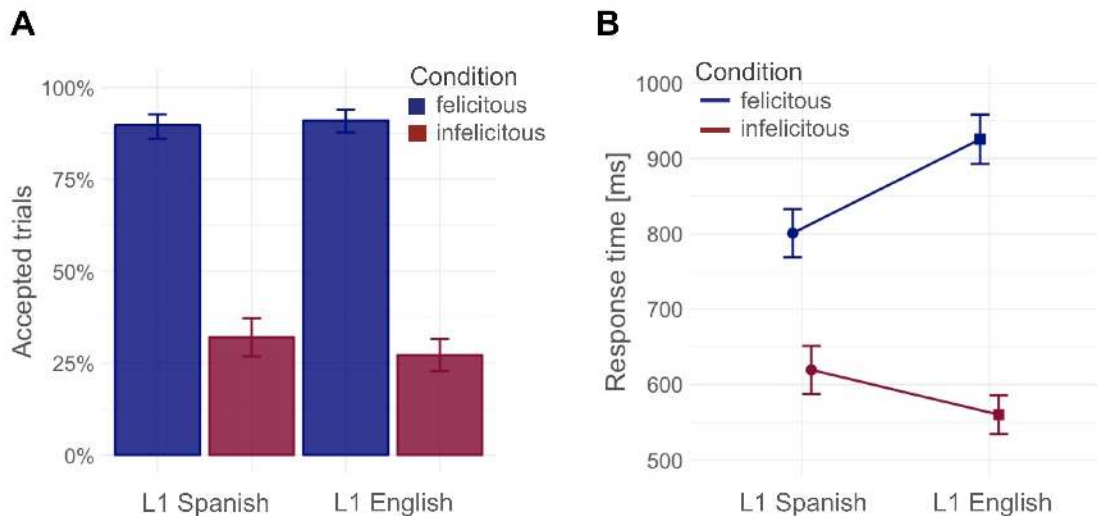
5.2 L1 Spanish group

Spanish speakers' responses showed a reliable effect of felicitousness: infelicitous sentences elicited fewer "acceptable" responses than felicitous sentences (32% vs. 90%) and also faster response times (619ms vs. 801ms). There was also an interaction between felicitousness and match, but none of the follow-up pairwise comparisons reached significance (*Felicitous sentences*: $\hat{\beta}_{Match}$ = -2.78; *SE* = 2.663; *z* = -1.063; *p* = .288; *Infelicitous sentences:* $\hat{\beta}_{Match}$ = 0.191; *SE* = 0.303; *z* = 0.629; *p* = .529). Participants' L2 English proficiency and the order of acquisition of their non-native languages did not significantly modulate their responses.

Table 2. Model results of Experiment 1. Model estimates ($\hat{\beta}$) are expressed in log odds for accuracy and log units for response times. For the FELICITOUSNESS factor, a negative estimate means that infelicitous sentences elicited fewer "acceptable" responses (or faster response times) than felicitous sentences. For the GROUP factor, a positive estimate means that Spanish speakers provided more "acceptable" responses (or showed longer response times) than English speakers. For the MATCH factor, a negative estimate means that mismatch conditions elicited fewer "acceptable" responses (or faster response times) than match conditions. For the PROFICIENCY factors, negative coefficients indicate that increasing levels of proficiency reduced the proportion of "acceptable" responses or elicited faster response times. Significant effects at the α = .05 level are bolded.

	Acceptability				Response times			
	$\hat{\beta}$	SE	z	p	$\hat{\beta}$	SE	t	p
Group analysis								
Felicitousness	**-3.778**	**0.392**	**-9.645**	**.000**	**-0.603**	**0.076**	**-7.953**	**.000**
Group	0.032	0.258	0.124	.901	0.090	0.167	0.540	.592
Felicitousness × Group	0.531	0.676	0.786	.432	**0.534**	**0.142**	**3.774**	**.000**
L1 Spanish								
Felicitousness	**-3.724**	**0.597**	**-6.234**	**.000**	**-0.346**	**0.089**	**-3.895**	**.001**
Match	-0.331	0.310	-1.067	.286	0.063	0.071	0.883	.382
German proficiency	-1.579	1.812	-0.871	.384	-0.175	1.147	-0.153	.880
L2 proficiency	1.829	2.584	0.708	.479	-0.998	1.635	-0.610	.549
Felicitousness × Match	**1.240**	**0.610**	**2.032**	**.042**	-0.090	0.127	-0.710	.478
Felicitousness × German proficiency	-2.311	4.792	-0.482	.630	-0.890	0.738	-1.206	.248
Felicitousness × L2 proficiency	-0.406	6.932	-0.059	.953	0.052	1.038	0.051	.960
L1 English								
Felicitousness	**-4.101**	**0.521**	**-7.868**	**.000**	**-0.867**	**0.106**	**-8.145**	**.000**
Match	-0.248	0.297	-0.833	.405	-0.137	0.067	-2.055	.048
German proficiency	-0.792	1.108	-0.715	.475	0.166	0.780	0.213	.834
L2 proficiency	0.187	0.835	0.224	.823	0.145	0.610	0.238	.814
Felicitousness × Match	0.444	0.553	0.803	.422	-0.063	0.119	-0.533	.594
Felicitousness × German proficiency	**-5.710**	**2.838**	**-2.012**	**.044**	-0.690	0.750	-0.920	.367
Felicitousness × L2 proficiency	-1.086	2.185	-0.497	.619	-0.862	0.575	-1.500	.148

Figure 1. L1 influence in Experiment 1. (A) *Acceptability*: both groups accepted infelicitous sentences less often than felicitous ones, with no differences between groups. Bars height represents mean proportions of "acceptable" responses and error bars show 95% confidence intervals. (B) *Response times*: infelicitous sentences were judged more quickly than felicitous sentences, but the difference was larger for English than Spanish speakers. Points represent mean response times in correctly answered trials and bars show standard errors.

5.3 L1 English group

English speakers' responses showed a reliable effect of felicitousness: infelicitous sentences elicited fewer "acceptable" responses than felicitous sentences (27% vs. 91%) and also faster response times (560ms vs. 925ms). There was also an interaction between felicitousness and participants' German proficiency, showing that participants were more accurate as their German proficiency increased. However, none of the follow-up pairwise comparisons reached significance (*Felicitous sentences*: $\hat{\beta}_{\text{German proficiency}} = 3.847$; *SE* = 7.998; $z = 0.481$; $p = 0.631$; *Infelicitous sentences*: $\hat{\beta}_{\text{German proficiency}} = -1.942$; *SE* = 2.940; $z = -0.661$; $p = .509$). Participants' L2 Spanish proficiency and the order of acquisition of their non-native languages did not significantly modulate their responses.

5.4 Post-hoc analysis

The removal of participants with lower than 75% accuracy in the untimed test had the unintended consequence of excluding all Spanish natives with low English L2 proficiency, leaving only highly-proficient speakers (Mean$_{\text{L2proficiency}}$ = 86%, Min = 75%; Max = 100%). We were concerned that the absence of an L2 proficiency effect was due to the fact that the data did not contain enough variation to support such analysis.

To address this concern, we performed a post-hoc analysis with all Spanish native speakers, regardless of their performance in the untimed test. This analysis showed a significant interaction between L2 proficiency and felicitousness ($\hat{\beta} = -10.724$; *SE* = 3.538; $z = -3.032$; $p = .002$). Follow-up comparisons showed that as participants' English

proficiency increased, they were less likely to accept infelicitous sentences containing possessor agreement violations ($\hat{\beta}$ = -7.550; SE = 2.808; z = -2.689; p = .007; Figure 2). No effect was found for felicitous sentences ($\hat{\beta}$ = 3.394; SE = 2.107; z = 1.611; p = .107).

For English natives, the absence of the effect was unlikely to result from reduced variation in the proficiency scale, which ranged from 15%–88% even after the removal of participants who failed the untimed test. The supplementary analysis revealed similar results to the main analysis, with no effect of participants' L2 proficiency on the acceptability of possessor violations ($\hat{\beta}$ = 0.122; SE = 2.072; z = 0.059; p = .953). Overall, participants were more likely to accept felicitous than infelicitous sentences, but this effect was not modulated by their Spanish proficiency.

Figure 2. L2 influence in Experiment 1. Effect of L2 proficiency on acceptability for La speakers, regardless of performance in the untimed test. For English natives, L2 proficiency did not modulate acceptability whereas Spanish participants were less likely to accept infelicitous sentences as their English proficiency increased. The x-axis shows L2 proficiency ratings on a 0–100% scale, with vertical black bars representing the number of participants at each point of the scale. Lines represent the model estimates of L2 proficiency effects after all other variables were taken into account, with ribbons showing the standard error of the estimates (Prado & Ullman, 2009). Points display by-subject acceptability model estimates by condition.
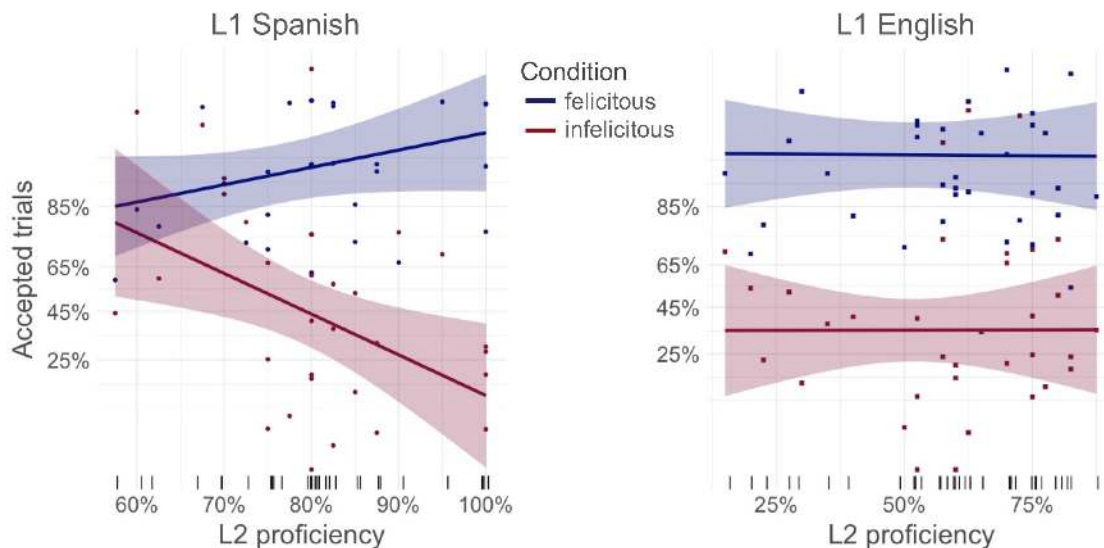
6 Discussion

Experiment 1 compared multilinguals' sensitivity to possessor agreement violations in German. As possessor violations are infelicitous in English and German, but not in Spanish, we were interested in whether multilinguals' judgments would be differentially modulated by their L1 grammar. The results of the judgment task showed between-group differences: although Spanish speakers accepted infelicitous sentences to similar extents than English natives, their response times showed a weaker effect of felicitousness, with

Spanish speakers taking longer than English speakers to reject infelicitous sentences. This suggests that Spanish natives, whose L1 grammar lacks possessor agreement, had more difficulty judging possessor violations in processing measures.

We also saw some evidence of L2 influence: in the post-hoc analysis, Spanish speakers with higher English proficiency were more likely to reject infelicitous sentences than speakers with lower English proficiency. The L2 effect was selective: Spanish speakers were better at rejecting possessor violations with increasing L2 English proficiency, but English speakers were not worse at detecting the same violations with increasing L2 Spanish proficiency. This result shows that high L2 proficiency may have helped Spanish speakers in German, but did not impair English speakers who knew Spanish, which lacks possessor agreement.

This outcome has two interesting implications. The first is that L2 influence seems facilitatory but not inhibitory, such that having been exposed to an L2 grammar may help, but not hurt, multilinguals' ability to process an additional non-native language. Second, the L2 influence on Spanish but not English speakers allows us to rule out a simplistic "more is better" account of foreign language learning. If multilinguals were good La comprehenders simply because they had learned two or more other languages, then facilitatory L2 effects should have been seen for both Spanish and English speakers, as both groups should have benefited from L2 exposure. Instead, our results suggest that an L2 plays a role specifically when the L2 and La grammars align.

Although the L2 effect is interesting and theoretically relevant, several limitations should be acknowledged. The first is that L2 influence was only observed when the whole Spanish group was analyzed: no L2 effect was seen with speakers with 75% accuracy or more in the untimed test. This fact suggests that Spanish speakers with lower English proficiency not only performed worse in the online task but were also unable to pass the untimed test, showing incomplete acquisition of German possessor constraints. A second limitation is that Spanish speakers were more proficient in their L2 than English speakers, and they had also acquired their L2 earlier. In fact, whereas most Spanish speakers had acquired English before German, approximately half of the English speakers had acquired Spanish and German simultaneously or in the reverse order. Although order of acquisition did not significantly modulate speakers' judgments, it is possible that their lack of an L2 effect was due to their demographic properties (including the order of acquisition of their non-native languages) rather than the fact that L2 influence cannot be inhibitory. Therefore, the proposal that L2 influence is helpful but not harmful should remain tentative due to these limitations, and further research with different language combinations and more L2-proficient groups is necessary.

Surprisingly, one factor that did not modulate our results was the gender match between the possessor and possessee nouns (i.e. we did not find more errors for *Mr. Schmidt... mother* than for *Ms. Schmidt... mother*). This was unexpected, as previous production studies had found that possessive errors were more frequent when the possessor and possessee noun mismatched in gender. If these results extended to comprehension, we would have expected Spanish speakers to wrongly accept infelicitous sentences (and to wrongly reject felicitous ones) more often in mismatch configurations. Spanish speakers' numerical patterns were consistent with this prediction (*Acceptability infelicitous sentences*: match = 29% vs. mismatch = 35%; *Acceptability felicitous sentences*: match = 93% vs. mismatch = 87%) but the effect of match did not reach significance. Therefore, although the lack of an effect cannot conclusively show that the effect is indeed absent,

these findings suggest that the mechanism for processing possessor agreement may differ between comprehension and production.

A final limitation of Experiment 1 is that speakers' knowledge of possessive constraints in their L2 was not independently measured. However, in order to argue that L2 influence was due to L2 knowledge, it is important to ensure that multilinguals indeed knew possessor constraints in their L2. This limitation was addressed in Experiment 2, which tested whether the results of Experiment 1 extended to a different comprehension paradigm in the absence of any metalinguistic judgments.

VI Experiment 2

Experiment 2 consisted of an implicit reading task, which allowed us to address whether the L1 and L2 effects in Experiment 1 arose due to the metalinguistic nature of the acceptability task, which required participants to judge the well-formedness of the sentences. In Experiment 2, participants answered comprehension questions, which were intended to focus their attention on interpretation rather than acceptability. Word-by-word reading times were used as a diagnostic of participants' implicit sensitivity to possessor violations in German: we hypothesized that if Spanish speakers had difficulty detecting possessor violations during reading, they would show reduced reading time disruptions after reading an infelicitous pronoun, as compared with English speakers. Further, we examined whether L2 proficiency additionally affected Spanish speakers' reading times, such that more English-proficient speakers showed stronger reading disruptions than less proficient speakers.

1 Participants

We recruited new groups of Spanish and English speakers of La German. The L1 Spanish group consisted of forty-seven Spanish natives with L2 English (mean age = 31 years, age range = 20–56, 34 females, 39 right-handed). The majority of Spanish natives had acquired English before German (n = 44). The L1 English group consisted of forty-three English natives with L2 Spanish (mean age = 32 years, age range = 16–65, 29 females, 36 right-handed). The majority of English natives had acquired Spanish before German (n = 33). The groups were matched in their German age of acquisition and proficiency (Table 1). Participants' Goethe scores were highly consistent with their German self-ratings (*L1 Spanish*: r = .70, t = 6.750, p = .000; *L1 English*: r = .60, t = 4.846, p = .000). Only speakers who scored 50% or more in the Goethe test were included in the analyses to ensure an intermediate-to-advanced level of German. Participants had acquired their L2 non-natively and had at most basic knowledge of other foreign languages.

2 Materials

Experimental materials consisted of 24 item sets, taken from the 32 experimental items in Experiment 1. Each sentence was followed by a yes/no comprehension question that probed its interpretation but did not target the pronoun-possessor relationship. For instance, for the item provided in (5), *Ms. Schmidt kissed her mother at the last family reunion*, the comprehension question was: *Did Ms. Schmidt miss the family reunion?* Experimental items were presented together with a set of 24 items from a separate experiment (half

ungrammatical) and 64 grammatical filler sentences of varied sentence structures. Across the experiment, half the trials had a target "yes" response and half had a target "no" response. Infelicitous trials represented 21% of the total number of trials. The experimental items were distributed across four lists in a Latin Square design and mixed with the other items and fillers. Randomization was performed on a by-participant basis.

3 Procedure

Participants were tested using a self-paced non-cumulative moving window design (Just, Carpenter & Woolley, 1982). The experiment was run in Ibex Farm (Drummond, 2013). Each trial began with a screen where the words were masked by dashes. The first time that participants pressed the space bar, the first word of the sentence was displayed. Each subsequent key-press revealed a word and re-masked the previous word (Figure 3A). The reading time for each region was measured as the time difference between two successive key presses. After each sentence, participants answered a yes/no comprehension question using the F ("yes") and J keys ("no"). Feedback was provided for incorrect responses. The experiment began by requesting participants to complete a demographic questionnaire. Then, they were instructed to read each sentence as quickly and carefully as possible. Participants completed three practice trials to familiarize themselves with the task.

As in Experiment 1, participants completed a series of additional tests: the Goethe Institute Placement Test, a vocabulary test, an untimed choice task on German possessor constraints and an untimed choice task on L2 possessor constraints. The L2 untimed test was either in English (L1 Spanish group) or in Spanish (L1 English group) and it consisted of 8 forced-choice trials. As in the German test, participants were instructed to respond based only on the context explicitly mentioned by the sentence. This is an example of an English item: *Jack asked _____ sister to be more patient* (options: *his* vs. *her*). Items with an unknown verb or possessee noun were excluded from analysis, and only participants with accuracy higher than 75% in the German test were included (all participants in this experiment), to ensure mature knowledge of German possessor constraints. An experimental session lasted approximately 70 minutes.

4 Analysis

Reading times for all trials were analyzed using linear mixed effects models. The regions of interest consisted of the pronoun (R4) and the three words following it (R5–7), because self-paced reading effects typically spill over after the critical word (Wagers et al., 2009; Parker & Phillips, 2016; Lago, Shalom, Sigman, Lau, & Phillips, 2015). To avoid analyzing each of the three spillover regions separately, which could increase the likelihood of Type 1 errors (von der Malsburg & Angele, 2016), the three regions were averaged on a trial-by-trial basis. The averaged response times were used as a summary of participants' behavior after encountering the pronoun.

The exclusion of items with unknown vocabulary affected 0.71% of the Spanish data and 0.68% of the English data. No participants were excluded due to accuracy lower than 60% in the filler trials. Following previous self-paced reading studies, reaction times faster than 200 ms or slower than 3000 ms were excluded. Exclusion percentages for the Spanish group were 0.45% (R4) and 0.63% (R5–7). Exclusion percentages for the English group

were 0.98% (R4) and 0.39% (R5–7). Following the Box-Cox procedure, performed on the reading times of regions R4 and R5–7, the reciprocal transformation was used.

The statistical analysis was performed in two stages. First, we addressed whether the two La groups differed from each other by using a model with FELICITOUSNESS, GROUP and their interaction as fixed effects. The second analysis considered each speaker group separately using the factors FELICITOUSNESS, MATCH in gender between possessor and possessee nouns, L2 and German proficiencies and the order of acquisition of non-native languages. Unless explicitly noted, the random effects structure of the models included uncorrelated by-participant and by-item intercepts and slopes for all categorical fixed effects.

5 Results

5.1 Group analysis

In the untimed task on German possessor constraints, all participants were highly accurate and accuracy did not differ between groups (*L1 Spanish*: 97%; *L1 English*: 97%, Wilcoxon signed-rank test, W = 1012, p = .986). Both groups were also highly accurate in the untimed task on L2 possessives (*L1 Spanish*: 98%; *L1 English*: 99%). In the self-paced reading task, speakers did not differ in their accuracy to the filler comprehension questions (*L1 Spanish*: 96%; *L1 English*: 96%; *ns. effect of Group*: $\hat{\beta}$ = -0.161; *SE* = 0.316; *z* = -0.510; *p* = .610) or experimental items (*L1 Spanish*: 95%; *L1 English*: 95%; *ns. effect of Group*: $\hat{\beta}$ = 0.147; *SE* = 0.363; *z* = 0.403; *p* = .687).

Figure 3B shows by-region reading times and Table 3 shows the results of the statistical analysis. No effects were found at the pronoun region. The post-pronoun regions showed a significant effect of felicitousness, with longer reading times in infelicitous than felicitous conditions. But crucially, there was an interaction between felicitousness and group: English speakers showed longer reading times for infelicitous than for felicitous conditions, but the difference was not significant for Spanish speakers.

5.2 L1 Spanish group

Spanish speakers did not show any effects in the pronoun region. In the post-pronoun regions, they showed numerically longer reading times for infelicitous than for felicitous conditions. However, the effect did not reach significance, suggesting a reduced sensitivity to possessor gender violations. There were no effects of L2 or German proficiency, of the order of acquisition of non-native languages or the gender match between possessor and possessee noun.

5.3 L1 English group

English speakers did not show any effects in the pronoun region. In the post-pronoun regions, they showed longer reading times for infelicitous than for felicitous conditions, suggesting that they were able to detect possessor violations. There was also an effect of their German proficiency, with speakers with higher German proficiency showing faster reading times than lower-proficiency speakers. This effect was more pronounced in felicitous ($\hat{\beta}$ = -1.649; *SE* = 0.515; *t* = -3.202; *p* = .003) than infelicitous conditions ($\hat{\beta}$ = -0.990; *SE* = 0.520; *t* = -1.904; *p* = .064). L2 proficiency, the order of acquisition of non-native languages and gender match did not modulate reading times.

Table 3. Model results of Experiment 2. Model estimates ($\hat{\beta}$) for the post-anaphor regions (R5–7) are expressed in reciprocal units. For the FELICITOUSNESS factor, a positive estimate means that infelicitous sentences elicited longer reading times than felicitous sentences. For the GROUP factor, a negative estimate means that Spanish speakers showed longer reading times than English speakers. For the MATCH factor, a positive estimate means that mismatch conditions elicited longer reading times than match conditions. For the PROFICIENCY factors, negative coefficients indicate that increasing levels of proficiency elicited faster reading times. Significant effects at the α = .05 level are bolded.

	Reading times			
	$\hat{\beta}$	SE	t	p
Group analysis				
Felicitousness	**0.106**	**0.025**	**4.189**	**.000**
Group	-0.034	0.109	-0.312	.755
Felicitousness × Group	**-0.091**	**0.044**	**-2.094**	**.039**
L1 Spanish				
Felicitousness	0.061	0.031	1.942	.065
Match	-0.015	0.033	-0.457	.652
German proficiency	-0.963	0.573	-1.681	.100
L2 proficiency	-1.153	0.676	-1.706	.095
Felicitousness × Match	0.059	0.054	1.105	.269
Felicitousness × German proficiency	0.077	0.217	0.353	.724
Felicitousness × L2 proficiency	-0.384	0.255	-1.502	.135
L1 English				
Felicitousness	**0.153**	**0.034**	**4.500**	**.000**
Match	0.011	0.036	0.315	.754
German proficiency	**-1.245**	**0.499**	**-2.496**	**.017**
L2 proficiency	-0.312	0.343	-0.908	.369
Felicitousness × Match	0.045	0.057	0.799	.425
Felicitousness × German proficiency	**0.639**	**0.228**	**2.809**	**.008**
Felicitousness × L2 proficiency	-0.130	0.157	-0.828	.413

Figure 3. L1 influence in Experiment 2. (A) *Self-paced reading task set-up*: words were presented one-by one, with each key press revealing a new word and hiding the previous one. (B) *Self-paced reading results (all regions)*: by-region means for experimental conditions. The post-pronoun regions are highlighted in gray. Error bars indicate the standard error of the mean. (C) *Self-paced reading results (regions of interest)*: condition averages in the post-pronoun regions: English speakers showed stronger sensitivity to infelicitous sentences than Spanish speakers.
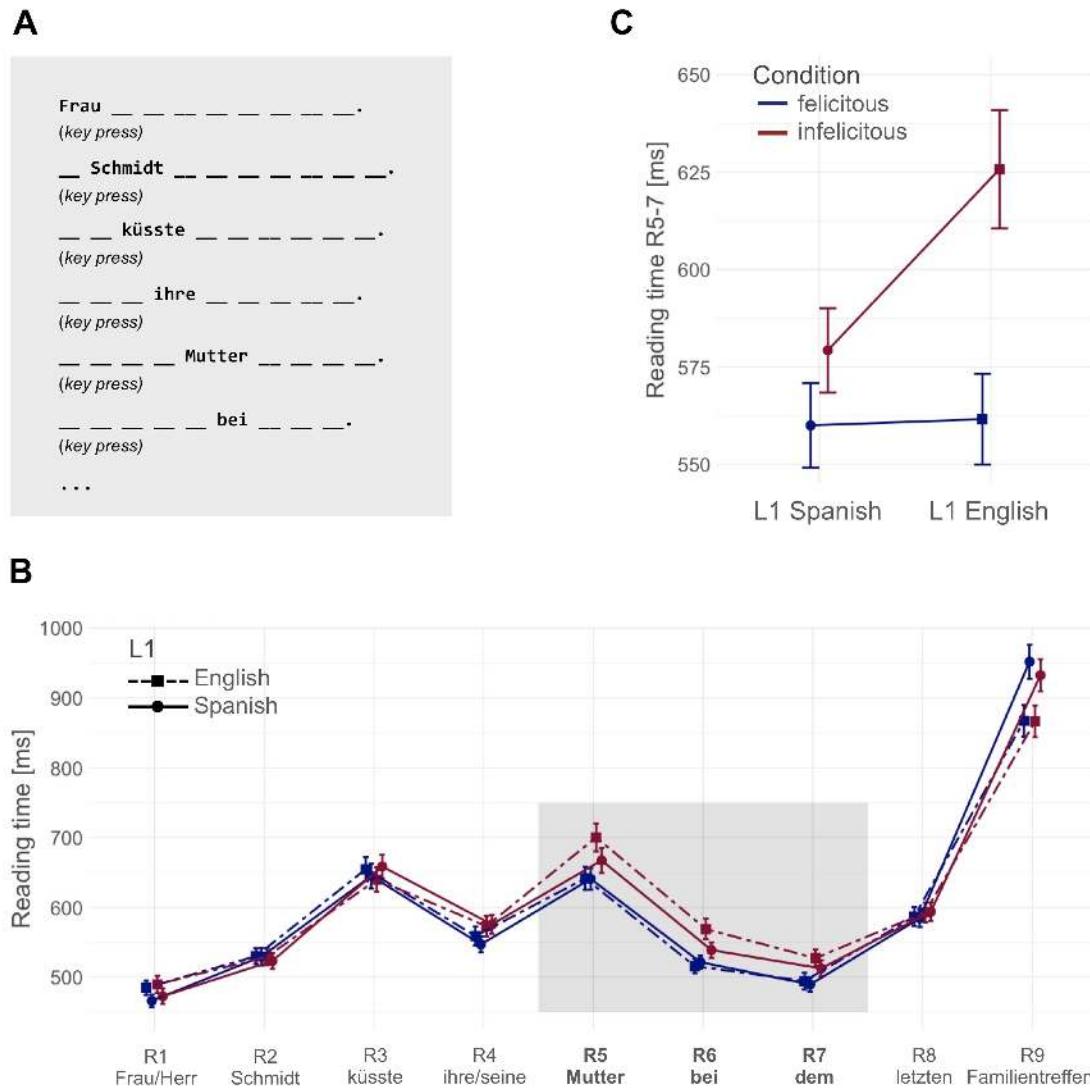
6 Discussion

Experiment 2 examined whether multilingual speakers were sensitive to possessor violations in a reading task without any metalinguistic judgments. After infelicitous pronouns, English natives showed larger reading disruptions than Spanish speakers. This result suggests that the fact that the English (but not the Spanish) grammar encodes possessor gender agreement rendered English speakers more sensitive to German gender-infelicitous pronouns during reading.

The second finding in Experiment 2 was that participants' L2 proficiency did not modulate their reading times. Notice that this result was unlikely due to participants' lack of awareness of L2 possessive constraints, as they were highly accurate in the untimed test that evaluated L2 pronoun knowledge. Finally, the gender match between the possessor and possessee noun did not modulate participants' processing, in contrast to previous production studies and consistently with Experiment 1.

VII General Discussion

Experiments 1 and 2 revealed a clear contrast between two groups of La German speakers, with L1 SPANISH-L2 ENGLISH speakers being less sensitive to German possessor agreement than L1 ENGLISH-L2 SPANISH speakers. In Experiment 1, both groups accepted infelicitous sentences to similar extents, but the response times of Spanish natives revealed a weaker effect of felicitousness than did the response times of English natives. In Experiment 2, Spanish natives showed reduced processing disruptions than English natives after reading gender-infelicitous pronouns. Lastly, some evidence of L2 influence was found in Experiment 1: Spanish natives with higher English proficiency rejected infelicitous sentences more often than less proficient speakers. In what follows, we consider the implications of these findings. Then, we turn to the mechanisms that might underlie the processing of possessive pronouns in multilingual speakers.

VIII Implications for La comprehension

This study investigated whether the comprehension of German possessor agreement was affected by the native Spanish and English grammars of multilingual speakers of German. Since the grammars of English and German encode possessor gender agreement but the grammar of Spanish does not, we hypothesized that Spanish native speakers would be less sensitive to gender-infelicitous pronouns than English native speakers. Additionally, we examined the potential effects of L2 influence and typological similarity. We hypothesized that if participants were influenced by their L2 agreement constraints, then their L2 proficiency might further modulate their sensitivity to possessor violations. Finally, if participants were influenced by the grammar typologically closest to German (English), regardless of its L1 or L2 status, then the groups might not differ because both should be able to deploy English agreement constraints.

The Spanish native speakers' response times (Experiment 1) as well as their online reading times (Experiment 2) supported the role of L1 influence. One reason for why the L1 grammar may have been so influential may be the nature of the tasks, which tested comprehension under processing pressure. Specifically, in the speeded acceptability task, words appeared quickly one-by-one, and participants provided their judgments under a response deadline. In the self-paced reading task, participants read at their own pace but

the words of the sentence appeared (and disappeared) one-by-one, thus forcing participants to rely on their working memory and preventing re-reading. This processing pressure, which also affects naturalistic language processing, might have encouraged reliance on L1 processing mechanisms, which may be more available and automatized than L2 mechanisms due to having been acquired earlier and used more frequently.

With regard to L2 influence, some support for this possibility was provided by the post-hoc analysis of the acceptability task: Participants with higher English proficiency were more accurate rejecting infelicitous sentences than participants with lower English proficiency. Interestingly, L2 influence was only facilitative: Spanish natives benefited from higher English proficiency but English natives did not become any less sensitive to possessor violations at increasing levels of Spanish proficiency. However, the lack of inhibitory L2 influence in the English group should be taken tentatively due to several limitations (see the discussion of Experiment 1) and further research will be needed to determine the reliability of this pattern. Finally, L2 influence was task-selective and only affected explicit acceptability judgments. This selectivity may indicate that multilinguals can always resort to their L1 mechanisms to process a La, but that L2 knowledge might only be available in situations that encourage the use of conscious linguistic knowledge, such as grammatical tests. This might occur because the procedures available in a L2 grammar are likely to be less automatized, so they might need to be consciously invoked in order to inform La processing.

The presence of L1 and L2 effects indicates that both grammars can affect multilinguals' comprehension of possessive pronouns. The L1 effect suggests that when an agreement constraint is absent in a native language, speakers have more difficulty deploying it in a La. This result is relevant not only for La but also for L2 processing work, which has produced divergent findings with regard to the role of L1 influence, which is sometimes absent or affects offline and online measures differentially (for examples, see Liszka & Roberts, 2013; Hopp, 2009; Roberts, Gullberg, & Indefrey, 2008). On the other hand, the L2 effect suggests that exposure to a L2 grammar can further modulate processing by increasing sensitivity to constraints that are present in both the L2 and La. These findings resemble previous findings from acquisition, where learners appear to capitalize on contrasts from both the L1 and L2 (e.g. Bérkes & Flynn, 2012; Flynn et al., 2004; Onishi, 2016; Rothman, 2010). Moreover, they show that advanced La speakers may still encounter difficulties during real time comprehension despite showing mastery of the relevant constraints in untimed tasks.

It is important to point out that cross-linguistic influence in acquisition and comprehension might be modulated by different variables, or by the same variables but to different extents. For example, typological similarity is likely to play an important role in learners' hypotheses about a new language: In the absence of other knowledge, learners should use the language they perceive as most similar to their La to make inferences about its grammar. But in real-time parsing, cross-linguistic influence might further rely on the accessibility of prior grammatical procedures in speakers' minds. Because comprehension requires speakers to deal with quickly unfolding input, multilinguals might depend on the linguistic procedures that are more automatized and can be deployed more readily, regardless of whether these procedures belong to a language typologically similar to the La.

Our results suggest that online comprehension studies provide new opportunities to research multilingualism. First, because they can contribute to the development of

psycholinguistic models on how the (worldwide-increasing) number of multilingual speakers comprehend sentences in real time, even after they have successfully acquired an La. Second, the acquisition of a new language will unavoidably be affected by speakers' ability to parse sentences in real time, thus rendering the study of comprehension informative for acquisition (Phillips & Ehrenhofer, 2015; Cabrelli Amaro, 2012). Thus, the use of online paradigms, which have already enriched the field of second language acquisition, are likely to provide valuable tools for researchers interested in multilingualism (e.g. Roberts, Gonzalez Alonso, Pliatsikas, & Rothman, 2016; Rothman, Alemán Bañón, & González Alonso, 2015).

IX   The comprehension of possessor agreement

Two mechanisms were previously proposed to explain the processing of possessor agreement in a non-native language. The first attributes possessive errors to a universal/L1-independent tendency to establish agreement locally, within the same syntactic unit (Pozzan & Antón-Méndez, 2017). This proposal is unlikely to account for our data because we obtained evidence of L1 influence, with Spanish natives being less sensitive to possessor agreement than English natives.

An alternative account, which does allow for L1 influence, proposes that Romance speakers transfer their L1 agreement procedures to English by computing gender agreement between pronouns and possessee nouns (Antón-Méndez, 2011). This account was supported by the presence of gender-match effects in production, where Romance speakers mostly made possessive errors in configurations where the possessor and possessee mismatched in gender. But gender-match effects did not occur in our experiments (or in the eye-tracking study of Pozzan & Antón-Méndez, 2017) and did not modulate the acceptability of infelicitous sentences (Experiment 1) or their reading times (Experiment 2). This suggests that the gender-match between the possessor and possessee noun might not play the same role in comprehension than in production.

To integrate our findings with previous work, we suggest that L1 influence with possessives arises from two separate mechanisms, which differentially affect production and comprehension. Specifically, the licensing of the possessive's gender in the sentence *Frau Schmidt küsste* <u>*ihren*</u> *Vater* ('Ms. Schmidt kissed <u>her.</u>masc father') requires Spanish speakers to compute (i) gender agreement with a preceding possessor noun, *Ms. Schmidt*; (ii) gender agreement with a following possessee noun, *father*. Due to the lack of possessor gender agreement in their L1, Spanish speakers may be less able to retrieve the gender of the possessor noun from working memory in order to use it to compute agreement, as compared with English speakers, whose L1 has possessor gender agreement.[3] Crucially, as both production and comprehension require the reactivation of the possessor noun to license the pronoun, errors should affect both modalities.

In contrast, errors due to the presence of a mismatching possessee noun may only affect production because previous work has shown that speakers jointly plan nouns and their modifiers before articulating noun phrases (Schriefers, 1993; Schiller & Caramaza, 2003). Due to this advanced planning, both the pronoun and possessee noun should be

---

[3] Note that this account assumes that antecedent features (e.g. the possessor noun) can be retrieved independently, rather than all antecedent features being reactivated as a whole (for discussion, see Antón-Méndez, 2010). This is because the possessor gender does not need to be licensed in Spanish, but number and person features do, and thus reactivation of some of the antecedent's features is needed, even in Spanish.

simultaneously co-activated in speakers' memory when a possessive is uttered: in this window, Spanish speakers might be misled by the gender of the possessee noun and wrongly assign it to the possessive, whereas the absence of this procedure in English should prevent English speakers from making errors. By contrast, in comprehension speakers have not yet encountered the possessee noun by the time they read the possessive, and thus gender-interference from the possessee noun is less likely to occur.

This account predicts that the gender of a possessee noun should be less relevant in comprehension than in production, or that it might only affect processing in cases in which the possessee noun is highly predictable and is already pre-activated prior to the occurrence of the possessive. Preactivation is unlikely to have occurred in our sentences, which did not contain any context to guide comprehenders' expectations about the identity of the possessee noun. Therefore, we suggest that Spanish speakers' errors in production arise from both difficulties in the retrieval of a possessor noun and misuse of L1 possessee agreement procedures, whereas only the former mechanism impacts comprehension. The use of paradigms with increased time resolution, such as eye-tracking and event-related potentials, is needed to distinguish the contribution of these two mechanisms during processing.

Another opportunity for future research concerns the processing of gender agreement between the possessive and the following possessee noun, which was not examined in this study. Although we previously characterized the relationship between possessive pronouns and possessor and possessee nouns in terms of gender agreement, these two types of agreement are likely different: the within-phrase agreement between the pronoun and possessee noun involves a clear morpho-syntactic dependency. However, possessor agreement in both English and German occurs across phrases and is also semantic in nature, with possessor violations resulting in not only a morphological conflict, but also in the failure of a presupposition about the gender of the possessor noun. Therefore, it would be interesting to examine whether the online processing profiles of possessee gender violations differ from those reported for possessor violations. This result would also be informative for the question of cross-linguistic influence, as possessee agreement is encoded in the grammar of Spanish but not English, raising the possibility that speakers of these languages may differ in their processing of German possessives.

X   Conclusion

This study addressed whether the effects of L1 and L2 grammars, which have been previously observed in metalinguistic tasks, extend to reading comprehension. We tested multilingual speakers of German with different L1/L2 backgrounds to examine whether cross-linguistic differences affected their performance in processing tasks. The results of speeded acceptability judgment and self-paced reading tasks revealed differences in how both groups processed German possessive pronouns. We propose that L1 effects reflect the automatic recruitment of native processing mechanisms, whereas L2 effects are restricted to the explicit deployment of metalinguistic knowledge. Our findings complement previous studies on La acquisition and they provide evidence that multilinguals' prior grammatical knowledge can affect how they read and comprehend La sentences in real time.

Acknowledgements

Declaration of conflicting interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding statement

References

Antón-Méndez, I. (2010). Whose? L2-English speakers' possessive pronoun gender errors. *Bilingualism: Language and Cognition*, *14*, 318–331.

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: the case of Germanic syntax. *Second Language Research*, *23*, 459–484.

Bardel, C., & Falk, Y. (2012). Behind the L2 Status Factor: A neurolinguistic framework for L3 research. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood*. Amsterdam, The Netherlands: John Benjamins, pp. 61–78.

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*, 4, 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 3, 255–278.

Berkes, É., & Flynn, S. (2012). Further evidence in support of the Cumulative-Enhancement Model. In J. Cabrelli Amaro, S. Flynn, & J. Rothman J (Eds.), *Third language acquisition in adulthood*. Amsterdam, The Netherlands: John Benjamins, pp. 143–164.

Blanche, P. & Merino, B. J. (1989). Self-Assessment of Foreign-Language Skills: Implications for Teachers and Researchers. *Language Learning*, 39, 3, 313–338.

Bohnacker, U. (2006). When Swedes begin to learn German: from V2 to V2. *Second Language Research*, 22, 4, 443–486.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, Series B (Methodological), 26, 211–252.

Cabrelli Amaro, J. (2012). L3 phonology: An understudied domain. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), Third language acquisition in adulthood, pp. 33–60. Philadelphia, PA: John Benjamins.

Chemla, E., Cummins, C., & Singh, R. (2016). Training and timing local scalar enrichments under global pragmatic pressures. *Journal of Semantics*, 34, 1, 107–126.

Chow, W. Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology, 5,* 630.

Delattre, P. (1965). *Comparing the phonetic features of English, French, German and Spanish. An interim report*. Heidelberg: Groos.

Delattre, P. (1969). *The general phonetic characteristics of languages*. Santa Barbara: University of California.

Van der Slik, F., Van Hout, R., & Schepens, J. (2017). The Role of Morphological Complexity in Predicting the Learnability of an Additional Language: The Case of La (Additional Language) Dutch. *Second Language Research*, https://doi.org/10.1177/0267658317691322.

de Bot, K., & Jaensch, C. (2015). What is special about L3 processing? *Bilingualism: Language and Cognition*, 18,2, 130–144. doi:10.1017/S1366728913000448.

Dewaele, J.-M. & Salomidou, L. (2017). Loving a partner in a foreign language. *Journal of Pragmatics*, 108, 116–130.

Dillon, B., Clifton, Jr., C., & Frazier, L. (2014) Pushed aside: Parentheticals and Processing. *Language, Cognition and Neuroscience*, 29, 4, 483–498.

Drenhaus, H., Saddy, D., & Frisch, S. (2005). Processing negative polarity items: When negation comes through the backdoor. In S. Kepser & M. Reis (Eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives* (pp. 145–165). Berlin: de Gruyter.

Enochson K, & Culbertson, J. (2015) Collecting Psycholinguistic Response Time Data Using Amazon Mechanical Turk. *PLOS ONE* 10, 3, doi: 10.1371/journal.pone.0116946.

Falk, Y., & Bardel, C. (2011). Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research*, *27*, 59–82.

Falk, Y., Lindqvist, C., & Bardel, C. (2015). The role of L1 explicit metalinguistic knowledge in L3 oral production at the initial state. *Bilingualism: Language and Cognition*, *18*, 227–235.

Flynn, S., Foley,C. & Vinnitskaya, I. (2004). The Cumulative-Enhancement Model for language acquisition: comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism*, 1, 1, 3–16 doi:10.1080/14790710408668175.

Foote, R. (2009). Transfer in L3 Acquisition: The role of Typology. In Leung, Y.K. (Ed.) *Third language acquisition and universal grammar*. Multilingual Matters, pp. 89–114.

Foucart, A., Santesteban, M., Branigan, H. P., & Pickering, M. J. (2011). The effect of L1 syntax on the agreement of L2 possessive structures: Evidence from eye-tracking. Poster presented at the *17th Meeting of the European Society for Cognitive Psychology* (ESCoP). Donostia-San Sebastián, Spain.

García Mayo, M.P., Ibarrola, A., & Liceras, J.M. (2005). Placeholders in the English Interlanguage of Bilingual (Basque/Spanish) Children. *Language Learning*, 55, 3, 445–489.

García Mayo, M. P., & Rothman, J. (2012). L3 morphosyntax in the generative tradition: from the initial state and beyond. In J. Cabrelli Amaro, S. Flynn, and J. Rothman

(Eds.), *Third Language Acquisition in Adulthood*. Amsterdam: John Benjamins, 9–32. doi: 10.1075/sibil.46.04pil.

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011) Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistic Compass*, 5, 509–524.

Goethe Institut (2010). German Placement Test. http://www.goethe.de/cgi-bin/einstufungstest/einstufungstest.pl (retrieved 01/01/2015).

González Alonso, J., & Rothman, J. (2017). Coming of age in L3 initial stages transfer models: Deriving developmental predictions and looking towards the future. *International Journal of Bilingualism*, 21, 6, 683–697. doi: 10.1177/1367006916649265.

González Alonso, J., Rothman, J., Berndt, D., Castro, T., & Westergaard, M., (2016). Broad scope and narrow focus: On the contemporary linguistic and psycholinguistic study of third language acquisition. *International Journal of Bilingualism,* 21, 6, 1–12.

Hermas, A. (2010). Language acquisition as computational resetting: verb movement in L3 initial state. *International Journal of Multilingualism*, *7*, 4, 343–362. doi:10.1080/14790718.2010.487941.

Hermas, A. (2014). Multilingual transfer: L1 morphosyntax in L3 English. *International Journal of Language Studies*, *8*, 2, 10–24.

Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition, 12*, 4, 463–483.

Hsien-jen Chin, D. (2009). Language transfer in the acquisition of semantic contrast in L3 Spanish. In Leung, Y.k (Ed.) *Third language acquisition and universal grammar*. Multilingual Matters, pp. 30–54.

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.

Jaensch, C. (2013). Third language acquisition: Where are we now*? Linguistic Approaches to Bilingualism*, 3, 1, 73–93. doi:10.1075/lab.3.1.04jae.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 3, 228–238.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R package version 2*.0–11. Retrieved from http://CRAN.R-project.org/package=lmerTest.

Lamers, M.J.S., Jansma, B.M., Hammer, A., & Münte, T.F. (2006). Neural correlates of semantic and syntactic processes in the comprehension of case marked pronouns: Evidence from German and Dutch. BMC Neuroscience, 7, 23, 1–12.

Lago, S., Shalom, D., Sigman, M., Lau, E., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, 82, 133–149.

Leung, Y.I., (2003). Failed Features versus Full Transfer Full Access in the Acquisition of a Third Language: Evidence from tense and agreement. *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference* (GASLA 2002), ed. Juana M. Liceras et al., 199-207. Somerville, MA: Cascadilla Proceedings Project.

Leung, Y.I., (2005). L2 vs. L3 initial state: A comparative study of the acquisition of French DPs by Vietnamese monolinguals and Cantonese–English bilinguals.

Bilingualism: Language and Cognition, 8, 1, 39–61. doi:10.1017/S1366728904002044.

Leung, Y.I., (2007) Second Language (L2) English and Third Language (L3) French Article Acquisition by Native Speakers of Cantonese. *International Journal of Multilingualism*, 4, 2, 117–149. doi:10.2167/ijm041.0.

Lighthown, P. M. & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching: Effects on second language learning. *Studies in Second Language Acquisition*, 12, 429–48.

Marian, V., Blumenfeld, H., & Kaushanskaya, M. (2007). e Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 4, 940–967.

Martens, M. (1988) Recognition and production of pronouns by francophone learners of English as a second language. Masters thesis, Concordia University.

Muñoz, C. (1994). A case of frequency-based markedness. *Atlantis,* 41, 165–177.

Na Ranong S. & Leung Y-KI. (2009). Null objects in L1 Thai-L2 English-L3 Chinese: An empiricist take on a theoretical problem. In: Leung Y-KI (ed.) *Third language acquisition and Universal Grammar.* Clevedon: Multilingual Matters, 162–91.

Nieuwland, M. S. (2014). "Who's he?" Event-related brain potentials and unbound pronouns. *Journal of Memory and Language*, 76, 1–28.

Onishi, H. (2016). The effect of L2 experience on L3 perception. *International Journal of Multilingualism*, 13, 4, 459–475.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34, 739–773.

Parker, D. & Phillips, C. (2016). Negative polarity illusions and the encoding of hierarchical representations in memory. *Cognition, 157,* 321–339.

Phillips, C., & Ehrenhofer, L. (2015). The role of processing in language acquisition. *Linguistic Approaches to Bilingualism*, 5, 4, 409–453.

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.

Prado, E.L., & Ullman, M.T. (2009). Can Imageability help up draw the line between storage and composition? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 4, 849–866.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Roberts, L., Gonzalez Alonso, J., Pliatsikas, C., & Rothman, J. (2016) Evidence from neurolinguistic methodologies: can it actually inform linguistic/language acquisition theories and translate to evidence-based applications? *Second Language Research*, 34, 1, 125–143. doi: 10.1177/0267658316644010.

Roberts, L., Gullberg, M. & Indefrey, P. (2008). On-line pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, 30, 3, 333–357.

Roberts, L. & Liszka, S. (2013). Processing tense/aspect-agreement violations on-line in the second language: a self-paced reading study with French and German L2 learners of English. *Second Language Research*, 29, 4, 413–439.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 5, 1–20.

Rothman, J. (2010). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. Second Language Research, 27(1), 107–127. http://doi.org/10.1177/0267658310386439.

Rothman, J. (2013). Cognitive economy, non-redundancy and typological primacy in L3 acquisition: evidence from initial stages of L3 romance. In S. Baauw, F., A.C. Dirjkoningen and M. Pinto (Eds.) *Romance Languages and Linguistic Theory* 2011. John Benjamins, Amsterdam, 217–248.

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: Language and Cognition*, 18, 2, 179–190.

Rothman, J., Alemán Bañón, J. & González Alonso, J. (2015) Neurolinguistic measures of typological effects in multilingual transfer: introducing an ERP methodology. *Frontiers in Psychology*, 6. 1087. doi: 10.3389/fpsyg.2015.01087.

Rothman, J., & Cabrelli Amaro, J. (2010). What variables condition syntactic transfer? A look at the L3 initial state. *Second Language Research,* 26, 2, 189–218.

Rothman, J., & Halloran, B. (2013). Formal Linguistic Approaches to L3/Ln Acquisition: A Focus on Morphosyntactic Transfer in Adult Multilingualism. *Annual Review of Applied Linguistics*, *33*, 51–67. doi:10.1017/S0267190513000032.

Sanz, C., Park, H., & Lado, B. (2015). A functional approach to cross-linguistic influence in ab initio L3 acquisition. *Bilingualism: Language and Cognition*, 18, 2, 236–251. http://doi.org/10.1017/S1366728914000285.

Santesteban, M., Foucart, A., Pickering, M.J., & Branigan, H.P. (2010). Is selection of possessive pronouns/adjectives in L2 affected by L1 syntax? Poster presented at the *16th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP)*. York, UK.

Schepens, J., Dijkstra, A., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*, 15,1, 157–166.

Schepens, J., Dijkstra, T., Grootjen, F., van Heuven, W.J.B. (2013) Cross-Language Distributions of High Frequency and Phonetically Similar Cognates. *PLoS ONE,* 8, 5, e63006.

Schiller, N.O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory & Language*, 48, 169–194.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 841–850.

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism*, 21, 6, 651–665. http://doi.org/10.1177/1367006916655413.

Spada, N., Lightbown, P., & White, J. (2005) The importance of form/meaning mappings in explicit form-focused instruction. In A. Housen and M. Pierrard (Eds.) *Investigations in Instructed Second Language Acquisition* (pp. 199–234). Berlin: Mouton de Gruyter.

Sprouse, J. (2011) A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavioral Research Methods*, 43, 155–167.

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition*, 114, 447–454.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the Declarative/Procedural model. *Bilingualism, Language and Cognition*, 4, 105–22.

von der Malsburg, T. and Angele, B. (2016). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, 94, 119–133.

Wagers, M., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206–237.

Wagers, M.W., Phillips, C. (2014). Going the distance: memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology*, 67, 1274–1304.

Westergaard, M., Mitrofanova, N., Mykhaylyk, R. & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism,* 21, 6, 666–682. https://doi.org/10.1177%2F1367006916648859.

White, J. (1998). Getting the learners' attention: A typographical input enhancement study. In C. Doughty and J. Williams (eds) *Focus on Form in Classroom Second Language Acquisition* (pp. 85–113). Cambridge: Cambridge University Press.

White, J., Muñoz, C., & Collins, L. (2007). The His/Her Challenge: Making Progress in a 'Regular' L2 Programme. *Language Awareness*, 16, 278–299.

Zobl, H. (1985) Grammar in search of input and intake. In S. Gass and C. Madden (eds) *Input in Second Language Acquisition* (pp. 329–344). Rowley, MA: Newbury House.

Appendix I. Results of L1 German group

This appendix reports the results of two experiments on German native speakers, which complement the results presented in the paper "*The role of native and non-native grammars in pronoun comprehension.*" The experiments on German native speakers aimed to ensure the efficacy of the experimental materials to evoke the experimental effects of interest.

Experiment 1: speeded acceptability judgment task

1 Participants

Participants consisted of forty German native speakers (mean age = 27 years, age range = 18–53, 22 females, 33 right-handed). In this and following experiments, participants provided informed consent and received financial compensation for their participation. All procedures were in accordance with the Declaration of Helsinki.

2 Materials, procedure and analysis

The materials, procedure and analysis were the same as reported in the paper for Experiment 1. Following the Box-Cox procedure (Box & Cox, 1964), response times were reciprocally transformed. No participants showed lower than 60% accuracy in the filler trials. The exclusion of responses shorter than 200 ms affected 13.67% of the data for the L1 German group.

3 Results

Accuracy in the filler trials was 93%. Responses to experimental items showed an effect of felicitousness, with infelicitous sentences being accepted less often than felicitous sentences (21% vs. 96%). Infelicitous sentences also elicited shorter response latencies than felicitous sentences (564ms vs. 576ms) but this difference did not reach significance. There were no other effects or interactions. The results of the statistical analysis are shown in Table 4.

Table 4. **Model results of Experiment 1**. Model estimates ($\hat{\beta}$) are expressed in log odds for accuracy and log (L1 English/L1 Spanish) and reciprocal units (L1 German) for response times. For the FELICITOUSNESS factor, a negative estimate means that infelicitous sentences elicited fewer "acceptable" responses (or faster response times) than felicitous sentences. For the MATCH factor, a negative estimate means that mismatch conditions elicited fewer "acceptable" responses (or faster response times) than match conditions. Significant effects at the $\alpha = .05$ level are bolded.

	Acceptability				Response times			
	$\hat{\beta}$	SE	z	p	$\hat{\beta}$	SE	t	p
L1 German								
Felicitousness	**-5.630**	**0.577**	**-9.765**	**.000**	-0.022	0.097	-0.224	.824
Match	0.182	0.350	0.520	.603	0.087	0.069	1.256	.219
Felicitousness × Match	-1.073	0.672	-1.597	.110	-0.003	0.108	-0.026	.980

Experiment 2: self-paced reading task

1 Participants

Participants consisted of forty-one German native speakers (mean age = 25 years, age range = 18–37, 34 females, 37 right-handed).

### 2 Materials, procedure and analysis

The materials, procedure and analysis were the same as reported in the paper for Experiment 2. No participants were excluded due to lower than 60% accuracy in the filler trials. Data exclusion percentages for the L1 German group were 1.62% (R4) and 0.41% (R5–7). Following the Box-Cox procedure, performed on the reading times of regions R4 and R5–7, the reciprocal transformation was used.

3 Results

German speakers were highly accurate in the comprehension questions of both filler and experimental items (97% and 96% respectively). In the experimental sentences, they showed longer reading times in infelicitous than in felicitous conditions at the pronoun ($\hat{\beta}$ = 0.087; SE = 0.040; t = 2.162; p = .037) and post-pronoun region (R5–7).[4] Thus, German natives showed clear signs of processing disruption when encountering a pronoun that mismatched in gender with the possessor noun (Figure 4).

---

[4] Due to non-convergence, by-item slopes were removed in the analysis of the post-pronoun regions.

Figure 4. **Self-paced reading results in Experiment 2.** By-region means for experimental conditions: at the pronoun (R4) and post-pronoun regions (R5–7), German speakers showed longer reading times in felicitous than in infelicitous conditions. The three post-pronoun regions are highlighted in gray. Error bars indicate the standard error of the mean.
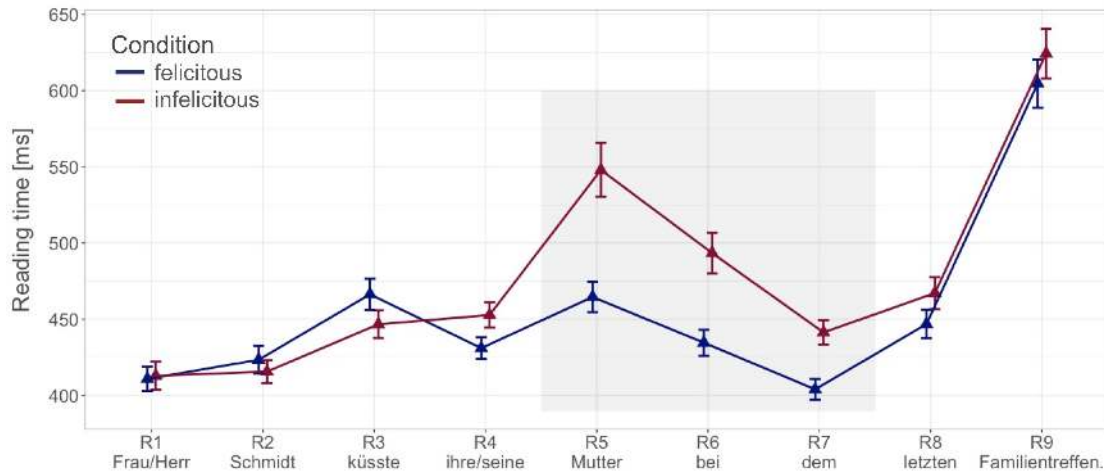
Table S3. **Model results of Experiment 2**. Model estimates ($\hat{\beta}$) for the post-anaphor regions (R5–7) are expressed in reciprocal units. For the FELICITOUSNESS factor, a positive estimate means that infelicitous sentences elicited longer reading times than felicitous sentences. For the MATCH factor, a positive estimate means that mismatch conditions elicited longer reading times than match conditions. Significant effects at the $\alpha = .05$ level are bolded.

	Reading times			
	$\hat{\beta}$	SE	t	p
L1 German				
Felicitousness	**0.188**	**0.044**	**4.293**	**.000**
Match	0.002	0.036	0.054	.957
Felicitousness × Match	-0.040	0.066	-0.608	.544