LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Possessive Processing in Bilingual Comprehension

Sol Lago,[a] Kate Stone,[b] Elise Oltrogge,[b] and João Veríssimo[c]

[a]Goethe University Frankfurt [b]University of Potsdam [c]University of Lisbon

Abstract: Second language (L2) learners make gender errors with possessive pronouns. In production, these errors are modulated by the gender match between the possessor and possessee noun. We examined whether this so-called match effect extends to L2 comprehension by attempting to replicate a recent study on gender predictions in first language (L1) German speakers (Stone, Veríssimo, et al., 2021). By comparing Spanish and English learners of L2 German whose languages have different possessive constraints, we were able to examine whether the match effect was modulated by the participants' L1. A first experiment suggested that predictions and match effects were absent in setups with complex visual displays. A second experiment with simpler

CRediT author statement – **Sol Lago**: conceptualization; methodology; data curation; writing – original draft preparation; writing – review & editing. **Kate Stone**: investigation; formal analysis; methodology; writing – review & editing. **Elise Oltrogge**: investigation; data curation; writing – review & editing. **João Veríssimo**: formal analysis; methodology; writing – review & editing.

A one-page Accessible Summary of this article in non-technical language is freely available in the Supporting Information online and at https://oasis-database.org

We thank José Alemán Bañon, Daniel Hole, Umesh Patil, and the members of the Knoeferle and Vasishth laboratories for useful feedback and discussion. We thank Lisa Becker, Eleonora Pinotti, and Janna Drummer for their assistance with material creation, participant recruitment, and data collection. Sol Lago and Kate Stone were supported by the German Research Foundation, project number 317308350. João Veríssimo was supported by grant UIDB/00214/2020 awarded to the Center of Linguistics of the University of Lisbon by the Foundation for Science and Technology, Portugal (FCT).

Correspondence concerning this article should be addressed to Sol Lago, Goethe University Frankfurt, Norbert-Wollheim-Platz 1, 60629 Frankfurt am Main, Germany. Email: sollago@em.uni-frankfurt.de

The handling editors for this manuscript were Aline Godfroid and Pavel Trofimovich.

DOI: 10.1111/lang.12556

displays successfully elicited predictions and match effects, but their size was comparable in Spanish and English speakers, inconsistent with crosslinguistic influence. We interpret our results as evidence that processing difficulties with possessives result from memory interference that impacts both L1 and L2 comprehenders.

Keywords prediction; gender agreement; crosslinguistic influence; sentence processing; visual world eye-tracking; German

Introduction

Possessive pronouns are notoriously hard to learn in a second language (L2). To date, these difficulties have been mostly demonstrated in spontaneous and elicited production in which even intermediate-to-advanced L2 English learners make gender errors, for example, *The dad put her little girl on his shoulder* (White, 1996).[1] Our study addressed whether possessive difficulties extend to L2 comprehension by using a processing effect recently found in first language (L1) German speakers (Stone, Veríssimo, et al., 2021). We examined the real-time mechanisms involved in possessive processing in order to arbitrate between two different explanations of L2 learning difficulties: Are they due to the misapplication of a L1 grammar? Or do they result from memory mechanisms that are shared by L1 speakers and L2 learners?

Background Literature

Previous Research on German Possessive Pronouns

German possessive pronouns are interesting for the study of real-time processing because they convey two different gender dependencies. Example 1 illustrates that, on the one hand, the stem of the possessive pronoun agrees in gender with a preceding possessor: *sein–* for masculine versus *ihr–* for feminine possessors (marked in bold in this and following examples). On the other hand, the suffix attached to the possessive pronoun agrees in gender with a following possessee noun: *–en* for masculine possessee versus –*e* for feminine possessee (marked through underlining in this and following examples).

Example 1

a. ***Elise***.$_{\text{POSSESSOR}}$ *vergaß bei dem Treffen* ***ihr****en Computer*.$_{\text{POSSESSEE}}$
"Elise forgot her.$_{\text{MASC.}}$ Computer.$_{\text{MASC}}$ at the meeting"

b. ***Dario***.$_{\text{POSSESSOR}}$ *vergaß bei dem Treffen* ***sein****e Brieftasche*.$_{\text{POSSESSEE}}$
"Dario forgot his.$_{\text{FEM.}}$ wallet.$_{\text{FEM}}$ at the meeting"

Due to the simultaneous encoding of possessor and possessee agreement, German comprehenders can implement at least two cognitive processes upon encountering a possessive pronoun: They can (a) identify a suitable antecedent (based on the gender of the pronoun stem) and (b) predict an upcoming noun (based on the gender of the pronoun suffix). These two agreement relationships are conceptually and syntactically different: Possessee agreement involves a within-phrase morphosyntactic dependency while possessor agreement involves a between-phrase dependency with semantic consequences. Thus, one might expect them to be computed independently.

However, a recent study demonstrated that possessor and possessee agreement interact during real-time processing (Stone, Veríssimo, et al., 2021). In two visual world studies, L1 German speakers looked at visual displays while hearing instructions like those in Example 2. The instructions contained a possessive pronoun that marked possessor and possessee agreement, for example, *Click on her*.$_{\text{MASC}}$ *button*.$_{\text{MASC}}$. The sentences varied in whether the possessor and possessee matched in gender. In the match condition, the possessor and possessee had the same gender, while in the mismatch condition they did not.

Example 2

a. Match condition (masculine target noun)
Klicke auf ***sein**<u>en</u>* *blauen* *<u>Knopf</u>!*
"Click on his.$_{\text{MASC}}$ blue.$_{\text{MASC}}$ button.$_{\text{MASC}}$"
b. Mismatch condition (masculine target noun)
Klicke auf ***ihr**<u>en</u>* *blauen* *<u>Knopf</u>!*
"Click on her.$_{\text{MASC}}$ blue.$_{\text{MASC}}$ button.$_{\text{MASC}}$"

The visual displays featured a button (the target noun, masculine in German) and a bottle (a competitor noun, feminine in German). Comprehenders' looks were used to assess whether the gender of the possessive suffix was used to predict the upcoming noun. The results showed that German speakers indeed used the possessive suffix predictively, thus showing a target-over-competitor looking preference shortly after the onset of the possessive. However, predictive looks to the target object occurred more quickly in the match than in the mismatch condition, suggesting gender interference from the pronoun stem. In the match condition in which the gender of the pronoun stem and suffix aligned, this gave comprehenders an advantage resulting in faster predictions. We refer to this pattern as a match effect.

Possessive Pronouns in a L2

One might deem the match effect in Stone, Veríssimo, et al. (2021) as likely to extend to L2 comprehenders because possessives often elicit errors in L2 speech. But in the L2 acquisition literature, possessive errors are typically attributed to a process different from gender interference: the misuse of L1 agreement mechanisms (Antón-Méndez, 2011; Lightbown & Spada, 1990; White, 1996). This is because possessive errors are frequently made by Romance speakers, whose L1 possessives establish forward-looking agreement with a following possessee noun, as shown in Example 3 for Spanish (Fabricius-Hansen et al., 2017; Lightbown & Spada, 1990; Muñoz, 1994; White, 1996, 1998; White & Ranta, 2002; White et al., 2007). Such a forward agreement mechanism is problematic if applied to a L2 like English, in which possessives establish only agreement with a preceding possessor noun, as shown in Example 4. Thus, an incorrectly gender-marked possessive as in *The dad put her little girl* suggests that Romance speakers establish agreement with the possessee, *the girl*, wrongly transferring their L1 agreement mechanisms to English.

Example 3

Nosotros queremos a nuestra hija.$_{\text{POSSESSEE}}$ *y a nuestro hijo*.$_{\text{POSSESSEE}}$
"We.$_{\text{MASC}}$ love our.$_{\text{FEM}}$ daughter and our.$_{\text{MASC}}$ son"

Example 4

Mr. *Müller*.$_{\text{POSSESSOR}}$ *loves* ***his*** *daughter and* ***his*** *son*

An elicited production study comparing Spanish, Italian, and Dutch proficient speakers of L2 English (Antón-Méndez, 2011) provided experimental evidence supporting the misapplication of L1 mechanisms. Spanish and Italian have possessee agreement whereas Dutch—like English—does not. The results showed that Romance speakers made more gender errors than did Dutch speakers: 15% versus 3% on average. Crucially, the errors were susceptible to a match effect, such that they decreased when the possessor and possessee had the same gender. For example, there were fewer errors with *His father got a new job* when the possessor was male. Italian and Spanish speakers produced similar amounts of errors, which is informative because, in contrast with Italian, the third person Spanish possessive *su* does not overtly mark gender agreement. Thus, the similar behavior of Italian and Spanish speakers suggested that Spanish, like Italian, has underlying syntactic agreement for third person possessive pronouns, only that the masculine and feminine possessors are identical in form (a conclusion further supported by the fact that possessee agreement in Spanish is still overtly realized with other grammatical features such as number (e.g., *su*.$_{SG}$ *hijo* vs. *sus*.$_{PL}$ *hijos*).

In contrast with L2 production, previous studies have repeatedly failed to find a match effect in L2 comprehension (Lago et al., 2019; Pozzan & Antón-Méndez, 2017). For example, Pozzan and Antón-Méndez (2017) tested 3–5-year-old English-speaking children and adult Mandarin Chinese speakers of L2 English. Both Mandarin speakers and English children produced more possessive errors when the possessor and possessee differed in gender, thus showing a match effect in production. However, a follow-up visual world experiment failed to elicit a match effect. Participants had to act out an instruction like *Give the apple to his sister* in a display with two adult characters (male and female), each with a younger male and female sibling. Eye movements were recorded to examine whether the possessor gender led participants to mistakenly predict a gender-matching possessee (e.g., a male sibling) as expected if they wrongly computed gender between the pronoun and the possessee noun. However, learners made almost no act-out errors, and the match between the possessor and possessee did not affect their eye movements in comprehension.

While the absence of a match effect in L2 comprehension was surprising, Pozzan and Antón-Méndez (2017) did not include comprehenders whose L1 encoded possessee agreement: Mandarin, in contrast with Romance languages, does not have possessee agreement. If the match effect is due to the misapplication of L1 mechanisms, then effects in comprehension may be more likely in speakers of L1s with possessee agreement (the General Discussion explains why Mandarin speakers may have shown a match effect in production). A later self-paced reading study that directly compared English and Spanish learners of German tested this possibility, but this study again failed to find a match effect (Lago et al., 2019). However, the use of a self-paced reading paradigm precluded the measurement of gender predictions as Pozzan and Antón-Méndez (2017) had done. This is because in the self-paced reading paradigm words are shown one at a time, and thus comprehenders do not know the identity of the possessee noun when the pronoun is encountered (by contrast, in the visual world task the possessee noun is pictured on screen, enabling its prediction).

Our study addressed this shortcoming in Lago et al.'s (2019) study by using a visual world paradigm to elicit gender predictions. We ensured that the paradigm was appropriate for detecting a match effect by replicating a design that had successfully elicited such an effect in L1 German speakers (Stone, Veríssimo, et al., 2021). Our study explored whether the match effect extended to L2 speakers of German. More importantly, we addressed whether the match effect was modulated by L1 influence by comparing two groups whose L1 grammars either had or lacked possessee agreement (Spanish vs. English). This comparison allowed us to gain insight into the source of difficulties with L2 possessives: Are they due to the misapplication of a native grammar? Or do they result from mechanisms shared by learners of different languages?

The use of a predictive paradigm additionally allowed us to address whether having an L1 with grammatical gender promotes the predictive use of gender in a L2. This has remained an unresolved question in L2 research (Foucart, 2021). To date, only one published eye-tracking study has directly compared the gender predictions of Italian (grammatical gender) and English (no grammatical gender) intermediate-to-advanced learners of Spanish (Dussias et al., 2013). However, the results were inconclusive with regard to L1 influence: Italian speakers showed predictive effects but only in trials with a feminine target noun. For English speakers, predictive effects depended on L2 proficiency: The most proficient speakers showed predictive effects while less proficient speakers did not. Our study revisited the role of L1 influence by comparing Spanish and English speakers of German.

The Present Study

We used the design and materials from Stone, Veríssimo, et al.'s (2021) study with intermediate-to-advanced Spanish and English speakers of German. Our predictions concerned two empirical effects and their susceptibility to L1 influence: the match effect and gender predictions. The match effect was relevant to examining whether the comprehension of possessives involved the misapplication of L1 mechanisms. Depending on the type of L1 influence, different outcomes might be expected. If Spanish speakers transfer their forward L1-possessee-agreement mechanism, they should be more likely than English speakers to wrongly compute agreement between the pronoun stem and the following noun, resulting in a larger match effect than for English speakers (as previously found in L2 production).

Alternatively, the lack of gender agreement in the stem of Spanish possessives might allow Spanish speakers to transfer their expectation that only the possessive suffix—rather than its stem—predicts the gender of the possessee noun. This is because possessive stems in Spanish only encode the person and number features of the possessor, for example, ***nuestr***$_{\mathbf{1ST.PL}}$-a$_{\text{.FEM}}$ *hija*$_{\text{.FEM}}$ versus **v*uestr***$_{\mathbf{2ND.PL}}$-a$_{\text{.FEM}}$ *hija*$_{\text{.FEM}}$ versus ***nuestr***$_{\mathbf{1ST.PL}}$-o$_{\text{.MASC}}$ *hijo*$_{\text{.MASC}}$ versus ***vuestr***$_{\mathbf{2ND.PL}}$-o$_{\text{.MASC}}$ *hijo*$_{\text{.MASC}}$. Due to this, Spanish speakers might transfer their expectation that only the possessive suffix predicts the gender features of the possessee, thus being more able to ignore the gender feature of the stem. If so, they should show a smaller match effect than would English speakers.

Crucially, both previous scenarios predict differences between Spanish and English speakers. Thus, they could be contrasted with an outcome of a similar match effect between groups. A similar match effect would be inconsistent with L1 influence and instead suggest a parsing mechanism for possessives that is shared across L2 learners.

The second effect of interest concerned gender predictions independent of the possessor–possessee match. We diagnosed gender predictions by measuring whether Spanish and English comprehenders differed in their ability to predict the target noun across conditions. Earlier predictions in Spanish versus English speakers would be consistent with L1 influence, suggesting that grammatical categories that are available in a L1 can be deployed more efficiently in L2 (Kaan & Grüter, 2021). Alternatively, predictions might be comparable between groups. This would suggest that having a gendered L1 does not, by itself, confer to learners an advantage using L2 gender predictively. Instead, predictions might depend more strongly on other variables, such as, for example, the congruency between gender systems, or they might be more influenced by individual-level variables such as L2 exposure, lexical access speed, and working memory (Dekydtspotter & Renaud, 2014; Hopp, 2013; Klassen, 2016; McDonald, 2006).

Experiment 1

We modeled the first experiment after Experiment 1 in Stone, Veríssimo, et al.'s (2021) study. L1 English and Spanish speakers heard auditory instructions like in Example 2 while looking at displays with four objects (see Figure 1A). The auditory instruction contained a possessive pronoun, a color adjective, and a noun indicating the target object. The target object matched the gender and color features in the instruction. The other objects matched only the color (color competitor), only the gender (gender competitor), or neither (distractor). The possessive indicated the gender of the target noun while the adjective indicated both its gender and color. The function of the adjective was to give listeners more time to process the gender features in the possessive. Given the differently colored objects on screen, the target object only became fully predictable at the adjective, which was the critical window for analysis. The crucial comparison was between the target and color competitor (henceforth, competitor): Because both objects matched the color of the adjective but only the target had the appropriate gender, any target-over-competitor advantage should have indexed the predictive use of gender information.

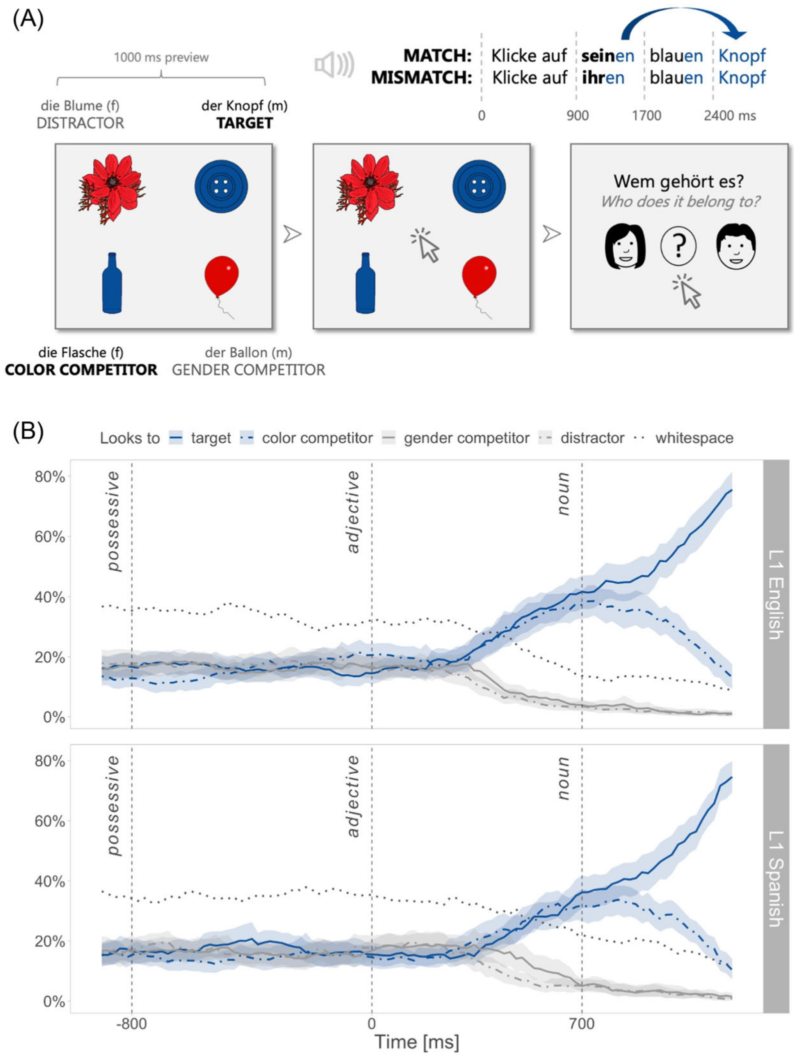

Figure 1 (A) Sample experimental trial in Experiment 1. (B) Percentage of fixations averaged across participants and items in Experiment 1. Horizontal lines show mean fixation percentages with 95% bootstrapped confidence intervals. Vertical lines show the onsets of the possessive, adjective, and target noun in the auditory instruction. Fixation curves start below 25% (chance level due to the four on-screen objects) because looks to whitespace are also plotted (dotted grey lines).

Table 1 Profiles of the English and Spanish participants in Experiments 1 and 2

Variables	Experiment 1		Experiment 2	
	L1 English	L1 Spanish	L1 English	L1 Spanish
Age of German acquisition (years)[a]	19 (10–31)	21 (5–34)	19 (10–32)	22 (5–44)
Goethe test score (out of 30)[b,c]	21 (4)	21 (5)	–	–
Self-rated German proficiency (%)[b]	72 (7)	73 (7)	73 (6)	74 (6)
Untimed possessive test accuracy (%)[b]	95 (7)	93 (9)	94 (8)	96 (7)

Note. Although some participants had an early age of exposure to German (e.g., 5–6 years), they were considered L2 learners because their exposure to German always occurred outside the home (e.g., in school) and because they self-identified as L2 speakers of German (for discussion, see Abrahamsson & Hyltenstam, 2009). [a]Mean age (age range). [b]Mean score (standard deviation). [c]Not applicable to Experiment 2.

To measure predictive effects, we evaluated both the size of the target-over-competitor advantage (time-window analysis) as well as the onset of this effect (onset analysis). The use of onset analyses allowed us to examine not only whether the participants were more likely to look at one object than the other but also when this preference emerged, enabling inferences about the onset of predictions (Stone, Lago, & Schad, 2021; Stone, Veríssimo, et al., 2021).

Method

Participants

Fifty-five L1 English speakers and 53 L1 Spanish speakers participated in Experiment 1. We matched the two groups in their age of acquisition and self-rated proficiency in German (see Table 1). We excluded data from participants who reported language impairments or exposure to languages other than their L1 at home during childhood. We also excluded participants if they had not rated themselves as having at least an intermediate level of German or if they had not achieved at least 70% accuracy in an untimed test probing for knowledge of German possessive pronouns (see Materials section). These two exclusion criteria ensured that our participants were intermediate-to-advanced German learners and that they had successfully acquired the grammatical construction under study. We entered 94 participants into the analysis: 47 L1 English speakers ($M_{\text{age}} = 32$ years, 28 female, 41 right-handed) and 47 L1 Spanish speakers ($M_{\text{age}} = 31$ years, 32 female, 44 right-handed). An ethics committee at the University of Potsdam approved this and the following experiments.

Materials

We took the materials from the first experiment reported by Stone, Veríssimo, et al. (2021). They consisted of 24 experimental items and 24 fillers. We controlled the 24 target nouns for their cognate status and gender congruency across languages. Specifically, all target nouns had the same gender in Spanish and German. Further, if the target nouns had cognates across languages (nine nouns), we ensured that the cognates were present in both Spanish and English. For example, *Batterie* in German, corresponding to *battery* in English and *batería* in Spanish. We spliced the auditory instructions such that the onsets of the possessive pronoun, adjective, and noun occurred at the same time across trials. A full list of materials with their identifiability ratings and gender distributions, together with the data and analysis code for all experiments, is publicly available online on IRIS (Lago et al., 2022a, 2022b, 2022c) and the OSF (https://osf.io/t6gnv). We used the 24 fillers to disguise that the purpose of the experiment was to study gender predictions. In the fillers, the participants could predict the target object based on the color of the adjective in the absence of a gender manipulation. We distributed experimental and filler trials into three lists[2]presented in a Latin square design such that each participant saw eight items per condition. We randomized the presentation order on a by-participant basis.

To ensure that a target-over-competitor advantage reflected the predictive use of gender rather than differences in visual saliency between objects, we made each object appear in the four roles across trials. For example, the color competitor in Example 2, *the bottle*, was the target object in another trial shown in Example 5 and the distractor and gender competitor in other trials. Across the experiment, half of the target nouns were masculine and half were feminine. For half of the trials, the possessive *sein* was used in the match condition and *ihr* in the mismatch condition. For the other half, the reverse was true. This was done to avoid a potential effect of referential ambiguity as the German possessive *ihr* can mean both *her* and *their*. The participants were told that the pronouns always referred to either a male or female character and the practice session was used to monitor that the participants resolved coreference as intended. We counterbalanced the position and color of the objects across trials.

Example 5

a. Match condition (feminine target noun)
   Klicke auf ***ihr****e gelbe Flasche!*
   "Click on her.$_{\text{FEM}}$ yellow.$_{\text{FEM}}$ bottle.$_{\text{FEM}}$"
b. Mismatch condition (feminine target noun)
   Klicke auf ***sein****e gelbe Flasche!*
   "Click on his.$_{\text{FEM}}$ yellow.$_{\text{FEM}}$ bottle.$_{\text{FEM}}$"

Procedure

The participants sat in front of a 21-inch computer monitor with a screen resolution of 1920 × 1080 pixels. Each object image was 300 × 300 pixels, such that the four images covered approximately 17% of the screen. The eye-to-screen distance was approximately 96 cm. Only the right eye was recorded at a sampling rate of 1000 Hz by an EyeLink1000 Plus eye tracker (SR Research, 2021). An adjustable chin and forehead rest served to minimize head movement during the experiment. At the beginning of the experiment, the participants were introduced to two characters, Martin and Sarah, whose faces were displayed on screen. The participants were asked to help Martin and Sarah find their belongings to tidy up their messy house before their parents arrived. They were told that they would see four objects while hearing an instruction and that they should click on the object mentioned in the instruction as quickly and accurately as possible. Each trial was followed by a question asking about the owner of the object (possessor identification task). This ensured that the participants resolved the coreferential relationship between the possessive and its antecedent (see Figure 1A). The results of this task were not of theoretical interest; thus, we have presented them in Appendix S2 in the Supporting Information online.

At the beginning of the session, the participants completed an object naming test with the nouns that would later appear in the experimental and filler trials. They were asked to name grey-colored objects using a determiner + noun sequence, for example, *der Knopf*. We conducted the naming task before the eye-tracking task to ensure that the participants' gender assignment reflected their own lexical knowledge rather than any learning effects from the eye-tracking task. If the participants provided the wrong gender or noun, the trial featuring that noun as a target object was excluded from analysis. This was done on the basis of previous research on gender predictions that has underscored the importance of dissociating gender assignment from gender agreement (Hopp, 2013; Hopp & Lemmerth, 2018). A stricter exclusion criterion would have been to also remove trials in which the wrong gender was provided for either the target or the competitor noun. However, because our design required each object to appear both as a target and a competitor across trials, using the stricter criterion would have doubled the proportion of excluded trials—to more than half of the data in Experiment 2. Therefore, we chose the more lenient exclusion criterion to maximize statistical power. Finally, we did not conduct a naming session in Spanish to measure the lexical entries activated by Spanish speakers. Thus, we could not rule out the possibility that between-group differences were affected by lexical issues. This is a limitation of our study.

After the eye-tracking session, we measured the participants' German proficiency via self-ratings and the Goethe Institute Placement Test (Goethe Institute, 2010). Self-ratings were collected by asking the participants to self-rate their proficiency on a scale from 1 ("least proficient") to 10 ("most proficient"). The range of scores in the Goethe test placed both participant groups between the B1 and C1 levels of the Common European Framework of Reference for Languages, which corresponds to an intermediate-to-advanced level. The self-ratings averaged across the four skills were significantly correlated with the Goethe scores ($r = .56$, $p < .05$). We also used an untimed task to determine that the participants knew the grammatical constraints associated with German possessive pronouns (see Appendix S1 in the Supporting Information online). A testing session lasted approximately 40 min.

Data Analysis

We preprocessed the raw data in DataViewer (SR Research, 2021), downsampled them to 50 Hz and exported them to the statistical software R (R Development Core Team, 2022). We excluded trials if they contained only blinks (events of pupil loss) or saccades (events with reduced visual information uptake). In the critical time window, this resulted in the exclusion of one trial in the L1 Spanish group. We also excluded trials if the participants clicked on the wrong object in the trial (L1 English: 0.27%; L1 Spanish: 0.67%) or if they did not provide the correct gender or label of the target noun in the naming test (L1 English: 22.66%; L1 Spanish: 24.58%). We performed two analyses on the preprocessed eye-tracking data: for the critical time window and for onset times.

Critical Time Window

The critical time window extended from the onset of the adjective to 200 ms after the onset of the noun to account for the time taken to program and launch eye movements (Hallett, 1986; Salverda et al., 2014). We compared fixations between the target and color competitor, for example, "blue button" versus "blue bottle." We operationalized the predictive effect as the target-over-competitor advantage in the critical time window.

We fit Bayesian generalized linear mixed-effects models with a logit link function to fixations within the critical time window using the package brms (Bürkner, 2017). Bayesian models are valuable because they combine prior information with evidence from the data in order to obtain a probability distribution over the plausible values of a parameter—the parameter's posterior distribution. Thus, an experimental effect can be quantified in terms of the likelihood of its possible values. This is more informative than a binary statement about whether the effect is significant because it puts the focus on determining an effect size and direction along with its uncertainty (Cumming, 2014; Kruschke & Liddell, 2018).

We coded fixations to the target and color competitor as 1 and 0, respectively. The model estimated (a) the target-over-competitor advantage (i.e., a predictive effect, intercept term), (b) the difference in the predictive effect between groups (L1 main effect), (c) the difference in the predictive effect between conditions (main effect of condition); and (d) the difference in the effect of the match/mismatch manipulation between groups (L1 × Condition interaction). We sum contrast coded the predictors condition and L1 to reflect our hypothesis that predictions would increase in the match versus mismatch condition and in the Spanish versus English group: match 0.5/mismatch −0.5, Spanish 0.5/English −0.5. Although the groups had comparable proficiencies, we controlled for any between-participant differences by adding the scaled and centered proficiency self-ratings as a further predictor (main effect of proficiency). Each model had a maximal random effects structure by participants and items.

We used weakly informative priors that constrained the models to psycholinguistically plausible parameter estimates, but that ensured that the priors would not outweigh the evidence provided by the data (Gelman, 2020; Gelman et al., 2008). The prior for the probability of predicting the target across conditions followed a normal distribution of N(0, 1) on the log odds scale. This indicated that the proportion of fixations to the target would fall with 95% prior probability between 12% and 88%. We used the same prior for all remaining predictors including random effects.

Bayes factors (BF) were used to quantify evidence for the effects (a–d) mentioned above. For each question, the null hypothesis (H_0) was that there was no difference between groups or conditions, and the alternative hypothesis (H_1) was that there was a difference. We computed Bayes factors using bridge sampling (Bennett, 1976; Gronau et al., 2017; Meng & Wong, 1996). This provided the ratio of the marginal likelihoods of H_1 and H_0. A BF_{10} of 1 suggests equivalent evidence for either hypothesis. In line with Lee and Wagenmakers's scale (2013)—adapted from Jeffreys's (1961) study—a BF_{10} above 1 indicates evidence for H_1: Values between 1 and 3 are considered inconclusive, values between 3 and 10 provide moderate, and values over 10 suggest strong evidence. Conversely, a BF_{10} below 1 indicates evidence for H_0: Values between

1 and 0.3 and are considered inconclusive, between 0.3 and 0.1 provide moderate, and below 0.1 suggest strong evidence. These categories are only a guide because Bayes factors should be interpreted in a continuous fashion. Readers interested in a frequentist analysis of the time-window data and the onset times (see below) can find them in Appendix S3 in the Supporting Information online. The results of the frequentist analysis were consistent with the Bayesian analyses.

Onset Times

This analysis established the onset of predictive effects, focusing on two questions: first, whether the onset times of the two groups differed with regard to predictive and/or match effects (see section Group Results), and second, whether there was an onset difference between the match and mismatch conditions within each L1 group (see section L1 English and L1 Spanish Results).

To calculate the onset of an effect and its temporal variability, we followed the Bayesian bootstrapping approach described in detail in Stone, Veríssimo, et al. (2021). Our priors about the onset of the predictive effect for the match versus mismatch onsets within each L1 group considered that an onset could only arise in the 700 ms time window between the onsets of the adjective and noun (adding 200 ms for saccade planning). For the onset of predictions in each condition, we specified a normal distribution with a 95% probability of the onset being between 200 and 900 ms: N(550, 175). We derived all priors for the effects between conditions and groups from these priors: The prior for the match–mismatch difference was a normal difference distribution of the individual condition priors: N(0, 247). For the onset of predictions in each L1 group, our prior was the mean of their individual match and mismatch priors: N(550, 124). The prior for the English–Spanish group difference was a normal difference distribution of the L1 group priors: N(0, 175). Finally, the prior for the match effect difference between the speaker groups was a normal distribution computed from the normal difference distributions of the match and mismatch priors within each group (i.e., a difference of differences): N(0, 350).

We approximated a likelihood function using onset time data generated by a bootstrap procedure (Stone, Lago, et al., 2021). The bootstrap involved a generalized linear model comparing fixations to the target versus competitor in each timebin. The predictors were condition and L1, and a (scaled and centered) nuisance predictor for self-rated proficiency. The posteriors for the match/mismatch conditions within each L1 group were the Gaussian product of the respective prior and likelihood distributions (Smith, 2011). The posteriors for the mean onset of predictive fixations across conditions for each L1 group were the mean of their respective match and mismatch posterior distributions. To quantify the degree of evidence for differences in onset times between groups and conditions (questions a–d), we computed Bayes factors via the Savage-Dickey method (Dickey & Lientz, 1970). Since Bayes factors are sensitive to the choice of prior, we also conducted sensitivity analyses to examine how different priors may have changed our conclusions. We conducted sensitivity analyses for all comparisons relevant to L1 influence.

Results

We report two sets of results. The main results concern onset times, which were previously shown to be affected by match effects in L1 speakers. However, we also report an analysis on the proportion of fixations to the target versus color competitor during the critical window as this analysis is often used for visual world data (Barr, 2008). The critical window spanned from the onset of the adjective until the onset of the noun—extended 200 ms to account for saccade planning. For each predictor of interest, we report a posterior distribution that reflects the probability of different effect sizes given the statistical model and the data. We summarize each posterior with its mean (*M*) and 95% credible interval (CrI). The latter represents the interval in which the true mean is estimated to lie with 95% probability.

Critical Time Window

The two groups fixated on the target and color competitor similarly during the adjective time window (see Figure 1B) and there was moderate evidence against predictive, proficiency, or L1 main effects. All other effects were inconclusive (see Table 2).

Onset Times: Group Results

First, we addressed whether L2 learners showed a target-over-competitor advantage prior to the onset of the target noun, that is, a predictive effect. The posterior onset estimates across conditions were inconclusive in both groups because the 95% credible intervals spanned onsets both before and after the onset of the target noun (shifted 200 ms): $M = 896$ ms, 95% CrI [835, 958] in English speakers; $M = 919$ ms, 95% CrI [874, 963] in Spanish speakers (see Figure 2A).

Second, we compared the two speaker groups directly to address whether they differed in the onset of the target-over-competitor advantage averaged across conditions. The estimate of the between-group onset mean difference

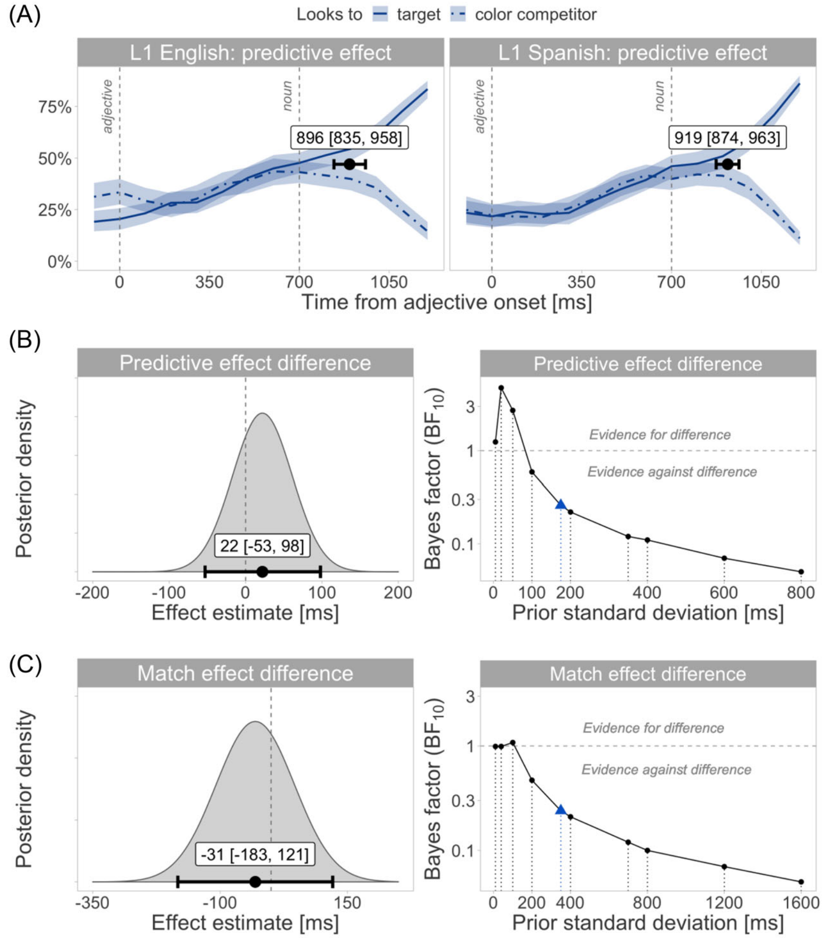

Figure 2 Onset time results in Experiment 1. (A) Estimated onset of the target-over-competitor advantage (the predictive effect) in each speaker group shown as a posterior mean and a 95% Bayesian credible interval. (B) Difference in the onset of the target-over competitor advantage between groups (left panel) and Bayes factor sensitivity analyses (right panel). (C) Difference in the onset of the match effect between groups (left) and Bayes factors sensitivity analyses (right). The priors used in the main analyses are shown with a blue triangle. The estimates of target-over-competitor advantage and the match effect were inconclusive.

Table 2 Results of the critical time window analysis in Experiments 1 and 2

	Experiment 1		Experiment 2	
Effect	*M* [95% CrI]	BF_{10}	*M* [95% CrI]	BF_{10}
Predictive effect (intercept)	0.07 [−0.27, 0.41]	0.19	0.39 [0.26, 0.53]	$> 5 \times 10^5$
Proficiency	0.03 [−0.19, 0.24]	0.11	0.08 [0.00, 0.16]	0.33
L1	−0.02 [−0.47, 0.44]	0.24	−0.15 [−0.38, 0.08]	0.28
Condition	0.19 [−0.27, 0.67]	0.40	0.11 [−0.05, 0.27]	0.16
L1 × Condition	0.34 [−0.53, 1.18]	0.72	−0.04 [−0.34, 0.25]	0.16
Gender congruency[a]	—	—	0.13 [−0.08, 0.34]	0.21
L1 × Congruency[a]	—	—	−0.04 [−0.39, 0.31]	0.13
Condition × Congruency[a]	—	—	0.12 [−0.19, 0.44]	0.19
L1 × Condition × Congruency[a]	—	—	−0.05 [−0.64, 0.53]	0.32

Note. Model estimates and 95% credible intervals (CrIs) are expressed in log odds. On the log odds scale, a value of 0 corresponds to chance level, with positive log odds indicating an advantage of the target over the competitor object. Bayes factors (BF) lower than 0.3 indicate evidence for the null hypothesis while Bayes factors above 3 indicate evidence for the alternative hypothesis. [a]Not applicable to Experiment 1.

was 22 ms, 95% CrI [−53, 98], and a Bayes factor of 0.26 indicated moderate evidence against a difference. The sensitivity analysis indicated moderate evidence for smaller effect sizes and increasingly strong evidence against larger effect sizes (see Figure 2B).

Finally, we addressed whether the Spanish and English groups differed in their match versus mismatch onset means (the magnitude of the match effect in each group is reported in the following sections). The posterior estimate of the mean difference in the match effect was −31 ms, 95% CrI [−183, 121], and a Bayes factor of 0.24 indicated moderate evidence against a group difference. The sensitivity analysis was inconclusive about smaller effect sizes but indicated strong evidence against larger effect sizes (see Figure 2C). Overall, these results did not support a predictive use of L2 gender and argued against between-group differences, either in the onset of the target-over-competitor advantage or in the size of the match effect.

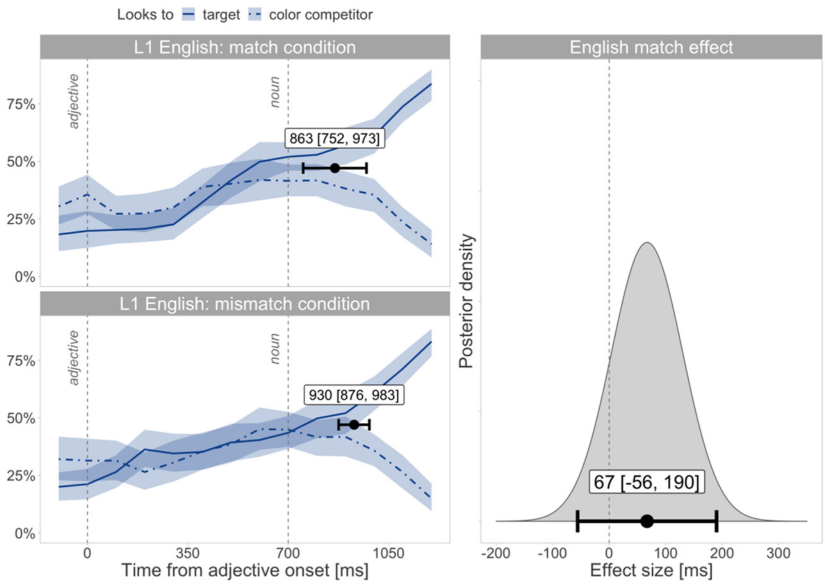

Figure 3 Onset time results in the English group in Experiment 1. Left panel: estimated onset of the target-over-competitor advantage by condition. Right panel: The estimated onset difference between the match and mismatch conditions was inconclusive.

Onset Times: L1 English Results

The onset of the target-over-competitor advantage had a mean of 863 ms, 95% CrI [752, 973], in the match condition and a mean of 930 ms, 95% CrI [876, 983], in the mismatch condition (see Figure 3). The mean difference between conditions was consistent with an earlier onset in the match condition, *Mdiff* = 67 ms, 95% CrI [−56, 190]. However, the effect was inconclusive, with its 95% credible interval spanning positive and negative effect sizes.

Onset Times: L1 Spanish Results

The onset of the target-over-competitor advantage had a mean of 901 ms, 95% CrI [826, 976], in the match condition and a mean of 936 ms, 95% CrI [888, 984], in the mismatch condition (see Figure 4). The mean difference between conditions was consistent with an earlier onset in the match condition, *Mdiff* = 36 ms, 95% CrI [−53, 125]. However, the effect was inconclusive, with its 95% credible interval spanning positive and negative effect sizes.

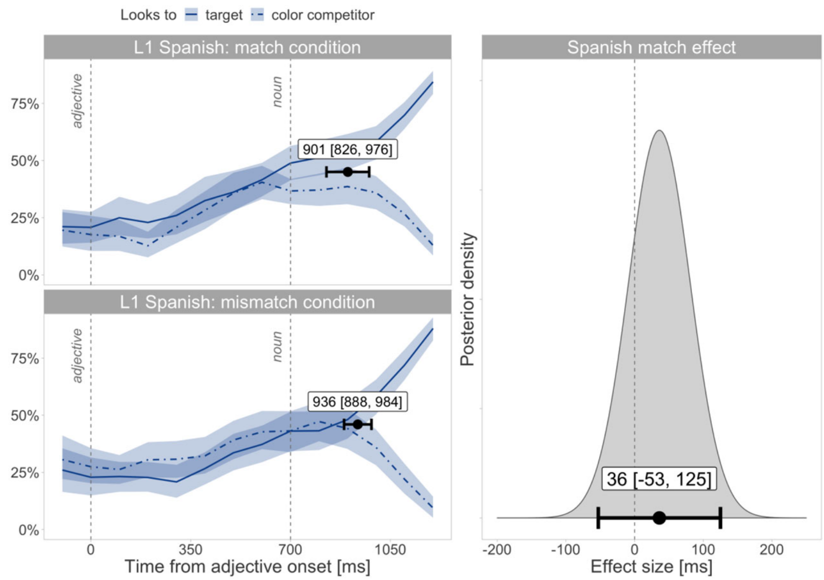

Figure 4 Onset time results in the Spanish group in Experiment 1. Left panel: estimated onset of the target-over-competitor advantage by condition. Right panel: The estimated onset difference between the match and mismatch conditions was inconclusive.

Exploratory Analyses

The main analyses of Experiment 1 included only trials with correct gender assignment in the naming test. However, a previous study by Hopp (2013) had shown that gender predictions may occur only in L2 speakers with a robust knowledge of gender across the whole set of nouns tested. Specifically, Hopp had found that participants with 5% or fewer errors in a naming test showed predictive effects but participants with 10% or more errors did not, even when he analyzed only the trials for which his participants had assigned the correct gender. To address whether a predictive effect emerged for our participants with robust gender knowledge, we reran our analyses considering only those participants who had made 5% or fewer errors in the naming test ($n = 10$, all learners with 100% naming accuracy). Consistent with Hopp's (2013) findings, the time-window analysis showed a predictive effect, with a target-over-competitor advantage during the adjective time-window. However, it was not possible for us to conduct an onset analysis to test the existence of a match effect because the model within the bootstrapping procedure failed to estimate an onset due to insufficient variability in the data. This was likely due to the limited number of learners (see Appendix S4 in the Supporting Information online).

Discussion

Experiment 1 compared English and Spanish speakers of German in their ability to predictively use gender-marked possessive pronouns. We examined whether we could replicate in L2 learners the match effect previously found in German L1 speakers (Stone, Veríssimo, et al., 2021). The between-group comparison allowed us to address two questions about L1 influence in L2 processing. First, we asked whether Spanish speakers would be better than English speakers at using gender predictively—consistent with an advantage due to their L1 having grammatical gender. Second, we examined whether Spanish speakers would be more sensitive than English speakers to the gender match between the possessive stem and suffix—consistent with the misapplication of L1 agreement procedures. The results, however, prevented us from answering these questions because we could not elicit gender predictions in either group. The time-window analysis did not show a target-over-competitor fixation preference in the critical time window. The onset analysis showed that the target-over-competitor preference arose only after the onset of the target noun.

Given the intermediate-to-advanced proficiency of our participants, it seems implausible that they could not use gender predictively. The absence of predictions was unlikely to have resulted from the participants' incomplete knowledge of the gender of the target words or of German possessives: We analyzed only trials in which the target objects (and their gender) had been named correctly and in which the participants had demonstrated accurate knowledge of German possessive pronouns in an offline task. Among the remaining variables that could explain the lack of predictions, one concerned our experimental design. We quantified predictive effects as a target-over-competitor looking preference, but previous studies have usually computed predictions as a difference in looks to a target object in a linguistically restricted versus nonrestricted contexts (the same/different paradigm; e.g., Grüter et al., 2012; Hopp, 2013).

A second possibility was that the absence of gender predictions was due to the specific setup of our visual world task. Two variables may have played a role. The first variable concerned the complexity of the visual displays that required the participants to recall the gender of four different nouns. This setup may have taxed their memory and reduced the resources available for predictive processing. In this regard, our study differed from some previous L2 studies on prediction that have featured simpler two-object displays (Dussias et al., 2013; Grüter et al., 2012; Lew-Williams & Fernald, 2010). While there have been previous studies with four-object displays, these involved the use of a single gender feature (e.g., a gender-marked determiner; Hopp & Lemmerth, 2018). In our study, the two different gender features of the possessive in combination with the grammatical gender of four different nouns may have made the task very demanding for L2 learners. A second variable was that the participants needed to use both gender and color features to predict the target nouns. Due to the visual nature of the task, the participants may have prioritized color over gender features. This tendency has been well documented in L1 speakers (Coco & Keller, 2015; Kukona et al., 2014). In L2 learners, the strategy may be even more prevalent because their gender representations may be less stable than are those of L1 speakers, thus causing them to rely more on visual than morphological cues.

To address these concerns, we conducted a follow-up experiment with a simplified experimental design. We modified our setup by presenting only two objects on the screen and by having both objects match in color but differ in gender. We speculated that a reduced number of objects on screen would be less taxing for the participants and free up resources for predictive processing. In addition, a two-object display would circumvent the need to integrate gender and color cues as only gender information was necessary to predict the target object.

Experiment 2

Experiment 2 addressed the same questions as did Experiment 1 but with a simplified two-object display in which only the target and color competitor were displayed on screen (see Figure 5A). Both objects had the same color but only the target matched the gender in the instructions. The color adjective was present to give the participants more time to process the gender cues of the possessive. The critical comparison was between fixations to the target versus the competitor, with a target-over-competitor advantage prior to the target noun reflecting the predictive use of gender information. The critical questions were whether this predictive effect would differ between groups and whether it would be modulated by the gender match between the possessive stem and suffix as a function of participants' L1.

We took the items for Experiment 2 from the second experiment reported by Stone, Veríssimo, et al. (2021). This experiment featured a larger number of experimental items; this allowed us to increase the number of trials in order to get more precise onset estimates. However, a potential disadvantage was that 18 of the new items included nouns that differed in gender between Spanish and German (e.g., the word "cage" is masculine in German, *der Käfig*, but

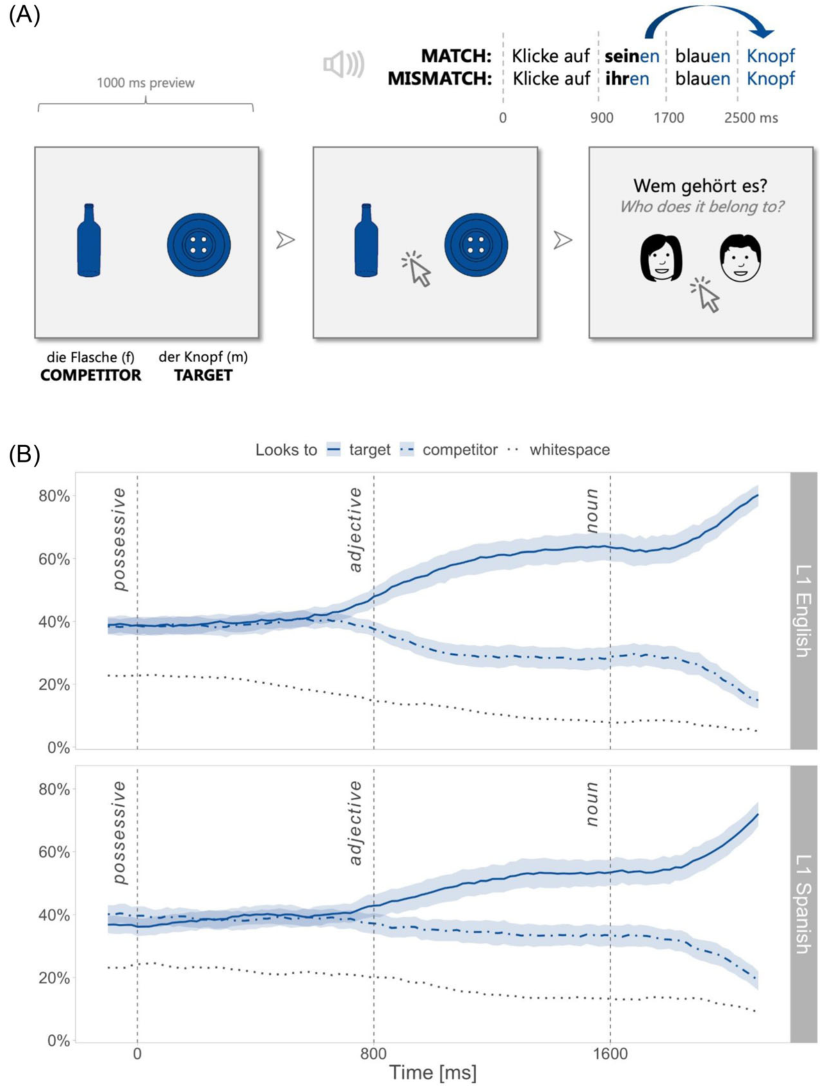

Figure 5 (A) Sample experimental trial in Experiment 2. (B) Percentage of fixations averaged across participants and items in Experiment 2. Horizontal lines show mean fixation percentages and shading shows 95% bootstrapped confidence intervals. Vertical lines show the onsets of the possessive, adjective and target noun in the instructions. Fixation curves start below 50% (chance level due to the 2 on-screen objects) because looks to whitespace are also plotted (dotted grey lines).

feminine in Spanish, *la jaula*). These trials could give rise to a crosslinguistic gender congruency effect: a delayed processing of L2 nouns when their L1 translations are of different genders in comparison to those with same gender (e.g., Costa et al., 2003). We again collected information about the participants' German proficiency after the eye-tracking session. In Experiment 2, however, the participants' proficiency was determined only through their self-ratings. We also administered the untimed test of knowledge of possessive constraints (see Appendix S1 in the Supporting Information online) that we had used in Experiment 1.

Method

Participants

Sixty-three L1 Spanish speakers and 69 L1 English speakers participated in Experiment 2. We excluded data from participants who reported language impairments or exposure to languages other than their L1 at home during childhood. As in Experiment 1, participants were also excluded if they had not rated themselves as having at least an intermediate level of German and if they had not achieved at least 70% accuracy in the untimed test on their knowledge of possessives. These two exclusion criteria ensured that the participants were intermediate-to-advanced German learners and that they had acquired the grammatical construction under study. This left 117 participants for the analysis (see Table 1): 63 L1 English speakers ($M_{\text{age}} = 30$ years, 37 female, 2 nonbinary, 56 right-handed) and 54 L1 Spanish speakers ($M_{\text{age}} = 34$ years, 43 female, 1 nonbinary, 53 right-handed).

Materials, Procedure, and Data Analysis

To increase the number of items, we added 72 new nouns to those used in Experiment 1. Of these, 60% had the same gender in German and Spanish, and 25% had a different gender. For the remaining 15%, gender congruency could not be unequivocally determined due to the availability of different Spanish terms for the same object; for example, the Spanish word for *stamp* can be either *sello* (masculine) or *estampilla* (feminine). The gender congruency of trials with these nouns was coded as NA (not applicable). The final materials consisted of 96 experimental items featuring two objects with the same color, resulting in 48 trials per condition. The onset of the target noun in the auditory instruction was 100 ms later than onset had been in Experiment 1.

The testing procedure was similar to that of Experiment 1, except that we used 18 filler trials featuring two objects with the same gender but different colors. We presented the fillers in a block after the experimental trials in order to remove the need for the participants to attend to color information during the experimental block. We used the fillers to check that the participants were able to generate predictions when they did not involve grammatical gender. Finally, we removed the "I don't know" option in the possessor identification task; participants had seldom used it in Experiment 1. Each object image was 300 × 300 pixels, such that the two images covered approximately 9% of the screen. The experimental session lasted approximately 1.5 hours.

For the analysis of the eye-tracking data, we excluded trials if they contained only blinks or saccades in the critical time window (there were none), if participants clicked on the wrong object in the trial (L1 English: 0.78%; L1 Spanish: 0.48%), or if the participants did not provide the correct gender or label of the target noun in the naming test (L1 English: 34.82%; L1 Spanish: 31.45%). In the time window analysis, the critical window ranged from possessive to noun onset, adding 200 ms for saccade planning. The critical window started at the possessive because the target object was already predictable at the possessive offset (in contrast with Experiment 1) and we were interested in the match/mismatch effect indexed by the possessive stem. We used the same weakly informative, regularizing priors as we had used for Experiment 1. The fixed effects in the model were the same as in Experiment 1, except that we added the sum contrast coded variable gender congruency to the analysis, together with its interaction with L1 and condition (congruent 0.5, incongruent −0.5). The onset analysis was also as in Experiment 1, except that we adjusted the priors to reflect the longer critical window. The results of the frequentist analyses (see Appendix S3 in the Supporting Information online) were consistent with those of the Bayesian analyses except that the main effect of proficiency in the critical time window—inconclusive in the Bayesian analyses—reached significance in the frequentist analysis. Since the Bayesian analysis was the planned method and the more conservative one, we considered the effect of proficiency to be inconclusive.

Results

Critical Time Window

In contrast with Experiment 1, there was a clear target-over-competitor advantage in the critical time-window, with strong evidence of a predictive effect (see Table 2 and Figure 5B). There was moderate evidence against main effects or interactions of proficiency, L1, or condition. There was also moderate evidence against any effect of gender congruency, suggesting that it did not influence fixation probability.

Onset Times: Group Results

L2 learners showed a clear target-over-competitor advantage, that is, a predictive effect, shortly after the possessive, $M = 724$ ms, 95% CrI [684, 763], in English speakers; $M = 798$ ms, 95% CrI [732, 864], in Spanish speakers (see Figure 6A and also Appendix S5 in the Supporting Information online). The estimate of the between-group mean difference in the predictive effect was 75 ms, 95% CrI [−2, 152], and a Bayes factor of 0.60 indicated inconclusive evidence against a difference (see Figure 6B). The sensitivity analysis indicated strong evidence for smaller effect sizes but was inconclusive about larger effect sizes.

The posterior estimate of the between-group mean difference in the match effect was 89 ms, 95% CrI [−65, 243], and a Bayes factor of 0.19 indicated moderate evidence against a difference (Figure 6C). The sensitivity analysis was inconclusive about smaller effect sizes but indicated increasingly strong evidence against larger effect sizes. Overall, these results supported a predictive use of L2 gender, but they argued against between-group differences either in the onset of the predictive or match effects.

Onset Times: L1 English Results

The onset of the target-over-competitor advantage had a mean of 652 ms, 95% CrI [584, 721], in the match condition and a mean of 795 ms, 95% CrI [754, 835], in the mismatch condition (see Figure 7). The mean difference between conditions supported an earlier onset in the match condition, $\underline{Mdiff} = 142$ ms, 95% CrI [63, 222]. A Bayes factor of 33 provided strong evidence of a match effect.

Onset Times: L1 Spanish Results

The onset of the target-over-competitor advantage had a mean of 683 ms, 95% CrI [574, 791], in the match condition and a mean of 914 ms, 95% CrI [839, 989], in the mismatch condition (see Figure 8). The mean difference between conditions supported an earlier onset in the match condition, $\underline{Mdiff} = 231$ ms, 95% CrI [99, 363]. A Bayes factor of 43 provided strong evidence of a match effect.

Exploratory Analyses

A concern with the onset results was that 25% of the target nouns differed in gender between Spanish and German. If these nouns created processing difficulties for Spanish speakers, this effect could have interacted with predictive and match effects, obscuring any potential advantage of the Spanish group over

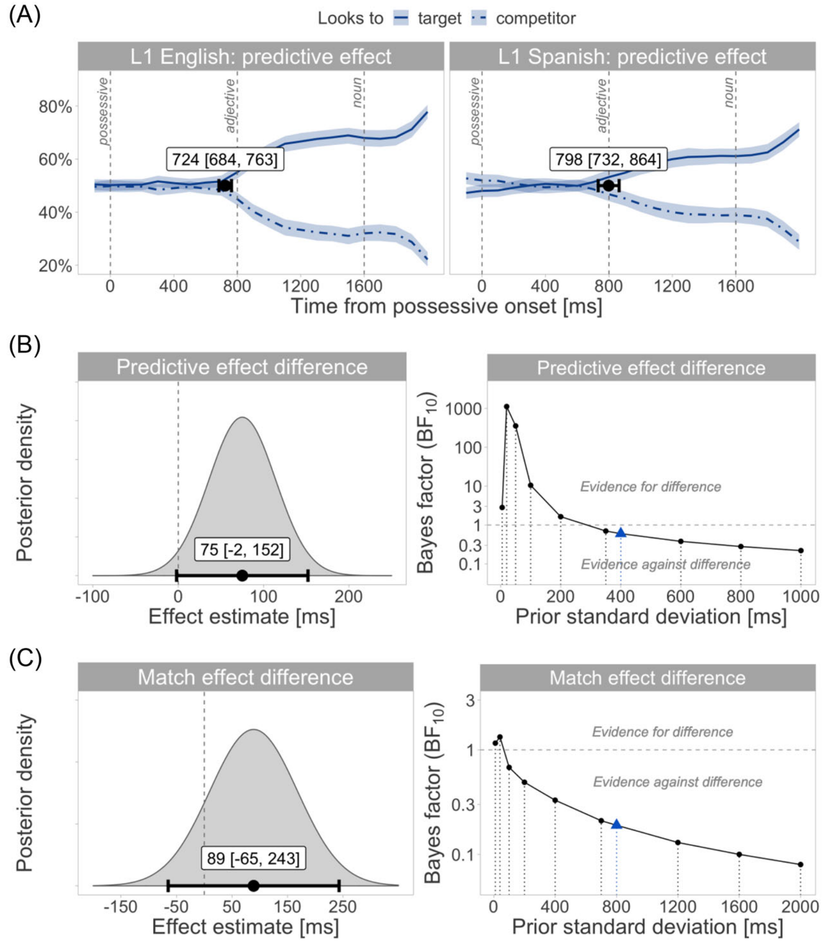

Figure 6 Onset time results in Experiment 2. (A) Estimated onset of the target-over-competitor advantage (the predictive effect) in each speaker group shown as a posterior mean and a 95% Bayesian credible interval. (B) Difference in the onset of the target-over competitor advantage between groups (left panel) and Bayes factor sensitivity analyses (right panel). (C) Difference in the onset of the match effect between groups (left) and Bayes factors sensitivity analyses (right). Priors used in the main analysis are shown with blue triangles. The analyses showed evidence of L2 gender predictions and match effects but not of between-group differences.

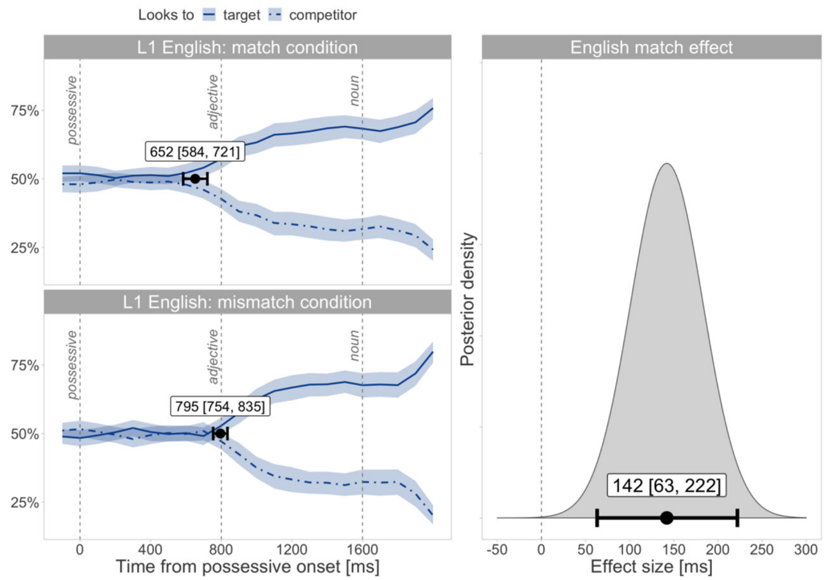

Figure 7 Onset time results in the English group in Experiment 2. Left panel: estimated onset of the target-over-competitor advantage by condition. Right panel: Estimated onset difference between the match and mismatch conditions. The estimate of the match effect supported an earlier onset in the match condition.

the English group. To address this concern, we performed a new analysis with only gender-congruent nouns, and we recomputed all between-group comparisons. These results were consistent with those of the main analyses. First, the estimate of the between-group mean difference in the predictive effect was 90 ms, 95% CrI [23, 157], and a Bayes factor of 2.72 indicated inconclusive evidence against a difference—the estimate suggested numerically earlier predictions in the English group, consistent with the tendency in the main analyses. Second, the posterior estimate of the between-group mean difference in the match effect was 20 ms, 95% CrI [−114, 154], and a Bayes factor of 0.08 indicated strong evidence against a difference.

Discussion

Experiment 2 used a simpler two-object display and a larger number of items to increase the likelihood of L2 predictions. All analyses found predictive effects in both groups. Further, the onset analyses revealed a match effect: Spanish and English speakers started predicting the target object earlier in the match

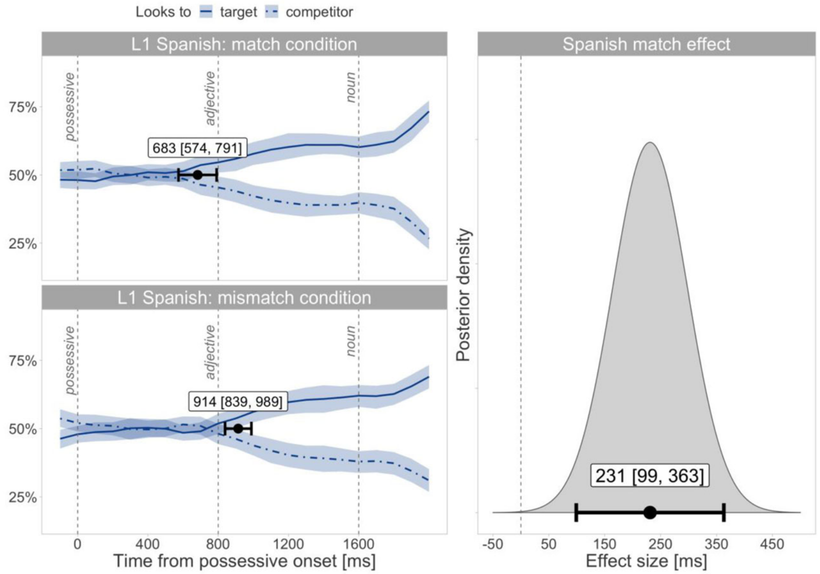

Figure 8 Onset time results in the Spanish group in Experiment 2. Left panel: estimated onset of the target-over-competitor advantage by condition. Right panel: Estimated onset difference between the match and mismatch conditions. The estimate of the match effect supported an earlier onset in the match condition.

condition, suggesting that the gender of the antecedent of the possessive interfered with their predictive mechanisms. Finally, the between-group comparisons showed that predictive and match effects were not stronger in the Spanish group than in the English group. These patterns persisted only when we analyzed gender congruent nouns to rule out a congruency effect.

General Discussion

Experiments 1 and 2 examined whether L2 speakers of German use the gender marking of possessive pronouns predictively and whether these predictions are modulated by the gender-match between the antecedent of the possessive and the target noun. Experiment 1 failed to elicit gender predictions, which prevented us from addressing whether they were modulated by a match effect. Experiment 2 used a simpler two-object display and more items and found strong evidence of predictive and match effects. These results with possessive pronouns extend previous findings with determiners and adjectives (Brouwer et al., 2017; Dahan et al., 2000; Dussias et al., 2013; Hopp, 2012, 2015; Hopp

& Lemmerth, 2018; Lew-Williams & Fernald, 2010). The results also show that the match between the possessive antecedent and the target noun shifts the onset of predictions in comprehension, thus conceptually replicating the match effect found in L1 German speakers (Stone, Veríssimo, et al., 2021). Finally, L1 influence did not modulate predictive or match effects. We discuss these two results separately.

Match Effects in Comprehension

In contrast with production studies on L2 possessives (Antón-Méndez, 2011; Pozzan & Antón-Méndez, 2017), previous comprehension studies had failed to find match effects (Lago et al., 2019; Pozzan & Antón-Méndez, 2017). Experiment 2 showed that L2 predictions are affected by a match effect but that this effect cannot be attributed to the misapplication of L1 agreement mechanisms. If so, the effect would have been weaker or absent in English speakers, who do not have forward possessive agreement. At first glance, our findings seem consistent with the explanation proposed by Pozzan and Antón-Méndez (2017), who suggested that difficulties with L2 possessives are due to a universal preference to establish agreement locally within the noun phrase. This bias was proposed as specific to production because this modality encourages the joint planning of the pronoun and possessee as a phrasal unit, thus promoting local agreement computations. We find a local bias account problematic for two reasons: first, because it states that the match effect should be restricted to production—contrary to what we found—and second, because the local bias was described as a default preference of language learners, which seems at odds with the finding that it affects adult speakers of L1 German—speakers who are not in the process of acquiring German (Stone, Veríssimo, et al., 2021).[3]

As an alternative explanation, we attribute the match effect to memory interference caused by the retrieval of the possessive antecedent. This explanation has been advanced for L1 speakers as a computational model within the cue-based retrieval theory (Patil & Lago, 2021). We propose that the same memory-based mechanisms operate in L2 speakers and thus that the match effect should be seen as resulting from processing limitations in working memory rather than a learning bias. Specifically, when comprehenders encounter a possessive, they first try to retrieve an antecedent using features in its stem as memory search cues (e.g., masculine). Because previously encountered entities are searched in parallel, they consider both nonlocal referents (e.g., Martin and Sarah) and local ones (e.g., the button and the bottle). Due to similarity-based interference, a local noun is sometimes misretrieved as the antecedent. This boosts its activation for prediction later on if the genders of the antecedent and target noun coincide (as in the match condition) or decreases it if the genders of the antecedent and target noun do not coincide (as in the mismatch condition). Thus, the match effect results from interference due to the co-activation in memory of the antecedent and target noun that disrupts both L1 and L2 processing.

Given our finding that possessive difficulties in comprehension are not specific to Romance speakers, an outstanding question is why Pozzan and Antón-Méndez (2017) did not observe a match effect with Chinese speakers of L2 English. One possibility is related to the setup of their visual world task, which was designed to be performed by children and may have been too easy for adults. Alternatively, the contrast between studies may reflect a crosslinguistic difference between English and German: German possessives encode two agreement dependencies (vs. one in English), which may render them very difficult to process, increasing the size—and thus the likelihood of experimentally eliciting—of the match effects. Further research is needed to arbitrate between these possibilities.

Bilingual Gender Predictions

The predictive effects that we found in Experiment 2 were not modulated by L1 influence. We had hypothesized that L1 Spanish speakers might be faster at using gender predictively in German due to the availability of gender in their L1. However, there was evidence against between-group differences, with predictions actually being numerically faster in English than Spanish speakers. These results suggest that L2 predictive abilities depend more strongly on variables other than that of whether learners' L1 encodes grammatical gender. While we controlled for between-group differences in proficiency and knowledge of the grammatical construction and target nouns, other variables may have played a role, for example, lexical access, processing speed, and/or working memory capacity (Bovolenta & Marsden, 2022). The effect of L1 influence may have been obscured by such variables, which suggests that they have a stronger influence than does a L1 on L2 predictive skills.

The contrast in our ability to measure predictions between Experiments 1 and 2 is surprising and has implications for future research. On the one hand, one could argue that two-object displays (such as the one in Experiment 2) are advantageous in a L2 setting because they reduce processing burdens associated with visual complexity and the need to combine different features to form predictions (e.g., color and gender in Experiment 1).[4] Ongoing research has suggested that these features may be relied on to different extents in L2 processing and that this depends on several variables, including task demands, listener goals, and the availability of other information (Henry et al., 2017; Henry et al., 2020; Kaan & Grüter, 2021). On the other hand, one limitation of the simpler, two-picture displays is that participants are more likely to become aware of the critical manipulation and start using the feature under study strategically (Curcic et al., 2019; Koch et al., 2021). This is challenging for interpreting experimental results in terms of the use of automatic processing mechanisms.

Limitations and Future Research Directions

Our study has limitations, and it also leaves some open questions for future research. First, it should be emphasized that our findings do not argue against any form of L1 influence. Our argument is more specifically that the match effect in comprehension cannot be exclusively due to the misapplication of L1 syntactic mechanisms to L2 processing. But other forms of L1 influence are possible. For example, previous work has shown that L2 learners are less effective at computing possessor agreement when the dependency is absent in their L1 (Alemán Bañón & Martin, 2021; Lago et al., 2019). This manner of L1 influence, that is, sensitivity to a dependency due to its L1 presence/absence, should be distinguished from the manner investigated here—L1 influence in terms of the misapplication of L1 syntactic rules to a L2.

Furthermore, as we explained in the Introduction, L1 effects could be hypothesized (a priori) to have opposing directions in L1 Spanish speakers: While the presence of forward L1 possessee gender agreement could increase match effects in German, the absence of possessor gender agreement could reduce them. Thus, if L1 influence had opposing effects across different participants—or across different trials of the same participant—these could have canceled each other out in the group average, preventing their detection.

A third limitation of our conclusions about L1 influence is that most of our Spanish participants had acquired English as a L2—its teaching is part of the primary school curriculum in most Spanish-speaking countries. Therefore, we cannot rule out that our Spanish speakers transferred processing routines from L2 English to German, obscuring potential differences between groups (e.g., Bardel & Falk, 2007; Rothman, 2011).

Finally, the picture that we have provided here with regard to possessive pronouns is necessarily incomplete, because the targets of prediction were feminine and masculine but not neuter nouns. For L1 processing, we would predict neuter nouns would elicit the same effects as do masculine or feminine nouns (later predictions when a masculine or feminine possessor mismatch a neuter possessee, for example *Klicke auf ihr*.NEUTER *Haus*.NEUTER). However, this prediction is complicated by the fact that there is no overt morphological marking for neuter possessee agreement in the configurations used in this study, thus, other syntactic configurations would be needed. Our predictions for L2 processing are less clear because they interact with architectural assumptions about how neuter gender is represented in learners whose L1 lacks a neuter category (e.g., L1 Spanish; for discussion see Klassen, 2016; Salamoura & Williams, 2007).

Conclusion

This study examined whether the comprehension of L2 possessives was affected by a match effect as had previously been found in L1 speakers. Our results demonstrated that this was indeed the case, such that match effects shifted the onset of L2 comprehenders' predictions. However, match effects were quantitatively similar between the two learner groups with different L1 possessive constraints, which is inconsistent with the misuse of L1 syntactic mechanisms as previously proposed for L2 production. Instead, we suggest that difficulties with possessives are due to memory interference during processing and that the cognitive mechanisms that create interference are common to L1 and L2 comprehenders.

Final revised version accepted 3 November 2022

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at https://osf.io/t6gnv and http://www.iris-database.org. All proprietary materials have been precisely identified in the manuscript.

Notes

1 We use the term possessive pronoun to highlight that forms like *his* or *her* require an antecedent to be interpreted. Syntactically, possessive pronouns function as modifiers of a head noun and inherit its morphosyntactic features. Depending on the language, possessives may behave like determiners or adjectives.

2 There were three lists because our experiment included an additional set of items with definite determiners instead of possessive pronouns, for example, *Klicke auf*

den blauen Knopf! Since L2 gender predictions have not previously been tested with possessive pronouns, we included the determiner items to confirm that our experimental design was able to replicate the predictive gender effects observed in previous research with articles and adjectives. This was indeed the case, and the results of the determiner items have been reported in Stone, Veríssimo, et al. (2021).

3 Note that these arguments are based on a literal interpretation of the claims of Pozzan and Antón-Méndez (2017). A reviewer argued that one could see our results as compatible in spirit, such that they would show (a) that a local agreement preference can affect comprehension when two contrasting morphological cues are simultaneously available during processing in production or comprehension and (b) that this preference/bias affects L1 speakers if the linguistic construction is complex enough.

4 Note, however, that the subset of L2 speakers with targetlike knowledge of grammatical gender could make predictions in contexts that manipulated both color and gender features as in Experiment 1 (see Appendix S4 in the Supporting Information online).

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Alemán Bañón, J., & Martin, C. (2021). The role of crosslinguistic differences in second language anticipatory processing: An event-related potentials study. *Neuropsychologia*, *155*, Article 107797. https://doi.org/10.1016/j.neuropsychologia.2021.107797

Antón-Méndez, I. (2011). Whose? L2-English speakers' possessive pronoun gender errors. *Bilingualism: Language and Cognition*, *14*(3), 318–331. https://doi.org/10.1017/S1366728910000325

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, *23*(4), 459–484. https://doi.org/10.1177/0267658307080557

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, *22*(2), 245–268. https://doi.org/10.1016/0021-9991(76)90078-4

Bovolenta, G., & Marsden, E. (2022). Prediction and error-based learning in L2 processing and acquisition: A conceptual review. *Studies in Second Language Acquisition*, *44*(5), 1384–1409. https://doi.org/10.1017/S0272263121000723

Brouwer, S., Sprenger, S., & Unsworth, S. (2017). Processing grammatical gender in Dutch: Evidence from eye movements. *Journal of Experimental Child Psychology*, *159*, 50–65. https://doi.org/10.1016/j.jecp.2017.01.007

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Coco, M. I., & Keller, F. (2015). The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *Quarterly Journal of Experimental Psychology*, *68*(1), 46–74. https://doi.org/10.1080/17470218.2014.936475

Costa, A., Kovacic, D., Franck, J., & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, *6*(3), 181–200. https://doi.org/10.1017/S1366728903001123

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Curcic, M., Andringa, S., & Kuiken, F. (2019). The role of awareness and cognitive aptitudes in L2 predictive language processing. *Language Learning*, *69*(S1), 42–71. https://doi.org/10.1111/lang.12321

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*(4), 465–480. https://doi.org/10.1006/jmla.1999.2688

Dekydtspotter, L., & Renaud, C. (2014). On second language processing and grammatical development: The parser in second language acquisition. *Linguistic Approaches to Bilingualism*, *4*(2), 131–165. https://doi.org/10.1075/lab.4.2.01dek

Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics*, *41*(1), 214–226. https://www.jstor.org/stable/2239734

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387. https://doi.org/10.1017/S0272263112000915

Fabricius-Hansen, C., Behrens, B., Pitz, A., & Helland, H. P. (Eds.). (2017). Possessives in L2 and translation: Basic principles and empirical findings. *Oslo Studies in Language*, *9*(2). https://journals.uio.no/osla/announcement/view/299

Foucart, A. (2021). Language prediction in second language. Does language similarity matter? In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 91–114). John Benjamins. https://benjamins.com/catalog/bpa.12.05fou

Gelman, A. (2020). Prior choice recommendations. Retrieved from https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wegenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97. https://doi.org/10.1016/j.jmp.2017.09.005

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215. https://doi.org/10.1177/0267658312437990

Hallett, P. E. (1986). Eye movements and human visual perception. In *Handbook of perception and human performance*. (Vol. *1*, pp. 10–11). Wiley.

Henry, N., Hopp, H., & Jackson, C. N. (2017). Cue additivity and adaptivity in predictive processing. *Language, Cognition and Neuroscience*, *32*(10), 1229–1249. https://doi.org/10.1080/23273798.2017.1327080

Henry, N., Jackson, C. N., & Hopp, H. (2020). Cue coalitions and additivity in predictive processing: The interaction between case and prosody in L2 German. *Second Language Research*, *38*(3), 397–422. https://doi.org/10.1177/0267658320963151

Hopp, H. (2012). The on-line integration of inflection in L2 processing: Predictive processing of German gender. In A. E. Kimball, A. K. Biller, & E. Y. Chung (Eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development* (pp. 226–245). Cascadilla.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, *53*(3), 277–306. https://doi.org/10.1515/iral-2015-0014

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 171–199. https://doi.org/10.1017/S0272263116000437

Jeffreys, H. (1961). *The theory of probability*. Clarendon.

Kaan, E., & Grüter, T. (2021). Prediction in second language processing and learning: Advances and directions. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 1–24). John Benjamins.

Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, *28*(1), 9–28. https://doi.org/10.1515/probus-2016-0002

Koch, E. M., Bulté, B., Housen, A., & Godfroid, A. (2021). Using verb morphology to predict subject number in L1 and L2 sentence processing: A visual-world eye-tracking experiment. *Journal of the European Second Language Association*, *5*(1), 115–132. https://doi.org/10.22599/jesla.79

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective.

Psychonomic Bulletin & Review, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4
Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 326–347. https://doi.org/10.1037/a0034903
Lago, S., Stutter Garcia, A., & Felser, C. (2019). The role of native and non-native grammars in the comprehension of possessive pronouns. *Second Language Research*, *35*(3), 319–349. https://doi.org/10.1177/0267658318770491
Lago, S., Stone, K., Oltrogge, E., & Veríssimo, J. (2022a). *Analysis code. Datasets from "Possessive processing in bilingual comprehension*" [Analysis code]. IRIS Database, University of York, UK. https://doi.org/10.48316/hr3d-bs88
Lago, S., Stone, K., Oltrogge, E., & Veríssimo, J. (2022b). *Data from "Possessive processing in bilingual comprehension"* [Dataset]. IRIS Database, University of York, UK. https://doi.org/10.48316/rrm3-g323
Lago, S., Stone, K., Oltrogge, E., & Veríssimo, J. (2022c). *Materials. Datasets from "Possessive processing in bilingual comprehension*" [Collection: Stimuli and experiment files]. IRIS Database, University of York, UK. https://doi.org/10.48316/6c4g-td51
Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003
Lightbown, 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*(4), 429–448. https://doi.org/10.1017/S0272263100009517
McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006
Meng, X.-L., & Wong, W. H. (1996). Simulating rations of normalizing constants vis simple identity: A theoretical exploration. *Statistica Sinica*, *6*(4), 831–860. https://www.jstor.org/stable/24306045
Muñoz, C. (1994). A case of frequency-based markedness. *Atlantis*, *16*(1/2), 165–177.
Patil, U., & Lago, S. (2021). Prediction advantage as retrieval interference: An ACT-R model of processing possessive pronouns. In T. C. Stewart (Ed.), *Proceedings of the 19th International Conference on Cognitive Modelling* (pp. 213–219). Applied Cognitive Science Lab, Penn State. http://acs.ist.psu.edu/papers/ICCM2021Proceedings.pdf

Pozzan, L., & Antón-Méndez, I. (2017). English possessive gender agreement in production and comprehension: Similarities and differences between young monolingual English learners and adult Mandarin–English second language learners. *Applied Psycholinguistics*, *38*(4), 985–1017. https://doi.org/10.1017/S0142716417000017

R Development Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.2) [Computer software]. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, *27*(1), 107–127. https://doi.org/10.1177/0267658310386439

Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, *10*(3), 257–275. https://doi.org/10.1017/S1366728907003069

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

Smith, J. O. (2011). Product of two Gaussian PDFs. In *Spectral audio signal processing* [Online book]. https://ccrma.stanford.edu/~jos/sasp/Product_Two_Gaussian_PDFs.html

SR Research (2021). *DataViewer* [Computer software]. https://www.sr-research.com/data-viewer

SR Research (2021). *Eyelink 1000 plus* [Apparatus and software]. https://www.sr-research.com/eyelink-1000-plus

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, *24*(5), 833–841. https://doi.org/10.1017/S1366728920000607

Stone, K., Veríssimo, J., Schad, D. J., Oltrogge, E., Vasishth, S., & Lago, S. (2021). The interaction of grammatically distinct agreement dependencies in predictive processing. *Language, Cognition and Neuroscience*, *36*(9), 1159–1179. https://doi.org/10.1080/23273798.2021.1921816

White, J. (1996). *An input enhancement study with ESL children: Effects of the acquisition of possessive determiners* [Unpublished doctoral dissertation]. McGill University. https://escholarship.mcgill.ca/concern/theses/z316q324x

White, J. (1998). Getting the learners' attention: A typographical input enhancement study. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 85–113). Cambridge University Press.

White, J., Muñoz, C., & Collins, L. (2007). The *his/her* challenge: Making progress in a 'regular' L2 programme. *Language Awareness*, *16*(4), 278–299. https://doi.org/10.2167/la419.0

White, J., & Ranta, L. (2002). Examining the interface between metalinguistic task performance and oral production in a second language. *Language Awareness*, *11*(4), 259–290. https://doi.org/10.1080/09658410208667060

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary
Appendix S1. Untimed Possessive Task.
Appendix S2. Possessor Identification Task.
Appendix S3. Frequentist Analysis of Experiments 1 and 2.
Appendix S4. Reanalysis of Experiment 1 Following Hopp (2013).
Appendix S5. Match Effect Across Second Language Groups in Experiment 2.