Check for updates

# Spanish−English bilingual heritage speakers processing of inanimate sentences

Rachel Casper[1]\*, Zenaida Aguirre-Muñoz[1,2], Michael Spivey[1] and Heather Bortfeld[1,3]

[1]Department of Cognitive & Information Sciences, University of California, Merced, Merced, CA, United States, [2]Department of Quantitative & Systems Biology, University of California, Merced, Merced, CA, United States, [3]Department of Psychology, University of California, Merced, Merced, CA, United States

This study investigated how heritage Spanish−English bilinguals integrate their cue hierarchies to process simple sentences in both Spanish and English. Sentence interpretation is achieved by weighing the various cues that are present in the sentence against that language's cue hierarchy. The Unified Competition Model (UCM) suggests that bilinguals show a variety of patterns in sentence interpretation strategies depending on language proficiency. Previous research on heritage Spanish−English bilinguals and late bilinguals has demonstrated differences in cue utilization and sentence interpretation compared to monolinguals. However, good-enough processing suggests that when a sentence does not meet certain heuristics, like the first-noun agent heuristic, a semantic representation of the sentence will be processed instead of a syntactic one. Even with reliable sentential-level cues such as word order, a plausible semantic representation of the sentence is favored. This is especially the case with inanimate−inanimate (IA-IA) sentences, like in the present study, in which there is less reversibility of thematic roles without competing with semantic plausibility. For this study, participants ($n = 32$) read a total of 80 inanimate sentences in English and Spanish with subject−verb−object and object−verb−subject (OVS) word orders, indicated the subject of the sentences, and completed language proficiency and dominance tasks. When reading Spanish sentences, participants read the OVS word order faster. English proficiency was a significant predictor of sentence reading time and choice selection time but did not predict word choice. The results suggest that IA-IA sentences pose challenges for cue utilization and thematic role assignment due to semantic limitations. This study found that participants may prioritize semantic plausibility over syntactic representations in sentence processing, supporting a good-enough processing model.

KEYWORDS

interpretation cues, good-enough processing, unified-competition model, heritage bilingual, animacy, semantic plausibility

## Introduction

When listening to or reading a sentence there are a few ways one can process the information to determine the lexical-syntactic roles and parse the sentence: lexical-based content of the sentence, extra-sentential content, or sentence internal content (Isabelli, 2021). For the present study, we describe lexical-based content and sentence internal content as they will be the most relevant due to the use of simple sentences. First, we discuss

sentence internal cues. Sentence internal interpretation cues can be used to determine the thematic roles of a sentence (MacWhinney, 2013). Thematic roles define the relationships between the verb and its nouns. Interpretation cues can have a variety of expressions across languages but largely fall into five categories: morphological case marking, animacy of nouns, intonation, verbal inflections, and syntax or word order (MacWhinney, 2013). The acquisition of these cues starts at an early age. Bates et al. (1984) found that children are sensitive to their language's cues from the age of 2. They suggest that it is the most important developmental achievement of sentence comprehension (Bates et al., 1984). Bilingual children acquire cues on a different timeline than monolingual peers (Reyes and Hernández, 2006), and bilinguals have different cue integration. Our study looks at how heritage speakers have integrated two of the most reliable cues from their two known languages, Spanish and English.

The Competition Model (CM) and the newer Unified Competition Model (UCM) suggest that sentence interpretation is achieved by weighing the various cues that are present in the sentence against that language's cue hierarchy (Bates and MacWhinney, 1989; Hernandez et al., 1994; Reyes and Hernández, 2006). According to Bates and MacWhinney (1989), adult monolingual English speakers rely most on word order followed by agreement and animacy. Adult monolingual Spanish speakers rely on differential object marking (DOM), which is touched on more in the Stimuli section, followed by verb agreement, clitic agreement, then word order (Bates and MacWhinney, 1989). When it comes to bilinguals, however, there are two competing cue sets (Bates and MacWhinney, 1989; MacWhinney, 2022) that vary based on language proficiency. Depending on the stage of acquisition and level of proficiency, bilinguals can have four different types of cue hierarchies (Reyes and Hernández, 2006): amalgamation (merged cues from all the known languages), differentiation (each language uses its own distinct cues), forward transfer (first-language, L1, cues are used on second-language, L2, stimuli), and backward transfer (L2 cues are used on L1 stimuli). Additionally, even though some bilinguals differentiate cues between their languages, there is still activation between the languages and their shared linguistic system (Reyes and Hernández, 2006).

In the past two decades, studies have suggested that Spanish–English heritage bilinguals use an amalgamation of the cue strategies from their languages (e.g., Hernandez et al., 1994; Reyes and Hernández, 2006). In both languages, word order and verb agreement are reliable cues, but the acquisition order and timeline of the cues in bilinguals are different than those of monolinguals (Hernandez et al., 1994; Reyes and Hernández, 2006). In a cross-sectional study, Reyes and Hernández (2006) found that bilinguals acquired the agreement cue earlier than English monolinguals but later than Spanish monolinguals. Their study on native (L1) Spanish, emergent English bilinguals examined the reaction times (RTs) and which noun the participants chose as the subject while they listened to sentences in English and Spanish in all six word order combinations. They found participants had an amalgamation and differentiation of cues from both languages (Reyes and Hernández, 2006). Furthermore, Nava (2007) found Spanish–English bilinguals used the subject–verb (SV) word order more than their Spanish monolingual counterparts when they

might have used a verb–subject (VS) word order to indicate known information, change of location, or an anticausative construction. In this study, bilinguals produced the subject-headed word order with the highest frequency. Nava concluded that the influence of English and the flexibility of word order in Spanish led to the SV construction preference. Together these studies suggest that Spanish–English bilinguals do not process or produce sentences in the same way as monolinguals in either of their languages.

Studies on heritage speakers of other languages also found differences among heritage bilinguals compared to monolinguals. In 2016, Pham and Ebert studied children with an average age of 7 in a Vietnamese–English transitional language program. In their longitudinal study, they found that over the 4 years, the participants shifted their use of cues. The children listened to animate and inanimate sentences in English and Vietnamese and slowly shifted toward relying on word order, a strong English cue, over animacy, the main cue in Vietnamese, over the duration of the study (Pham and Ebert, 2016). The researchers concluded that bilinguals process sentential information differently than monolinguals of either language even at a young age. Testing a similar age group, Meir et al. (2020) analyzed eye-tracking data from Russian–Hebrew bilinguals, Hebrew monolinguals, and Russian monolinguals. Children looked at images and were prompted to give a response to elicit the accusative case (a grammatical case used in some languages to indicate the direct object of a verb or the object of certain prepositions). Participants completed a visual world task, and their eye movements were recorded as they listened to sentences in both languages. In this study, the bilingual children were slower at processing the case cue than Russian monolinguals (Meir et al., 2020). For the Hebrew stimuli, upon hearing the beginning of an OVS sentence, specifically hearing the accusative case attached to the object, bilinguals began anticipating a subject (Meir et al., 2020). The Hebrew monolingual participants, by comparison, were not able to use the case cues predictively, although they used case in the production aspect of the experiment. The authors suggest that the strong Russian case cue is informing the weaker Hebrew case cue in bilingual sentence processing, creating a cue hierarchy and processing style different from that of monolinguals.

In the past year, two studies similar to that of Meir et al. (2020) have been published on Turkish heritage speakers in different countries. Özsoy et al. (2023) had monolingual Turkish and Turkish–German heritage speakers listen to sentences in Turkish while looking at a visual world paradigm display with reference objects. Subsequently, participants had to decide if a video showed the event described by the auditory stimuli. Heritage speakers do use case cues predictively; however, more monolinguals use predictive processing than bilinguals (Özsoy et al., 2023). The authors suggested this was possibly due to individual differences and learner background. Replicating the original same study, Karaca et al. (2024) tested Turkish monolinguals and Turkish–Dutch heritage speakers on verb-medial and verb-final sentences to see how the position of the verb affected cue predicting. The monolingual participants were able to predict the second noun phrase in the sentence upon hearing the case cue attached to the first noun phrase. Heritage-speaking participants, however, were only able to use case predictively in the verb-medial condition when

the verb semantics helped scaffold cue use (Karaca et al., 2024). Additionally, because bilingual participants' language experiences, especially literacy, in both languages were significant predictors, the authors lend support to prediction-by-production accounts and suggest that reading and writing training in either of the speaker's languages can help with cue prediction (Karaca et al., 2024).

Also looking at case cue processing, Chrabaszcz et al. (2022) studied adult heritage speakers of Russian residing in two different countries, Estonia and the United States, who therefore spoke two different L2s, Estonian and English. This study aimed explore how heritage speakers with two different dominant languages processed Russian cases. Participants listened to Russian sentences with a locative or instrumental case and selected an image that best represented the auditory stimulus (Chrabaszcz et al., 2022). They found that English-dominant, but not Estonian-dominant, speakers used word-order cues and misread Russian thematic cues in the locative and instrumental cases (Chrabaszcz et al., 2022). Estonian-dominant participants had native-like comprehension of the cases but slower task RTs (Chrabaszcz et al., 2022). They concluded these differences between the heritage-speaker groups were due to the lack of familiarity of English-dominant speakers with case morphology because of a dominant language with a different morpho-syntactic system (Chrabaszcz et al., 2022). In heritage bilinguals across languages, bilinguals process stimuli in both languages differently than monolinguals and have sentential processing that is deeply affected by the shared language system.

Studies on late Spanish–English bilinguals find differences in bilingual sentence parsing, analyzing a sentence for its constituents and their relationships, compared to that of native monolinguals. Isabelli (2021) found that fifth-semester Spanish learners were more accurate with OVS sentences when there was context to constrain their interpretation. These late bilinguals used semantic and pragmatic cues to avoid reliance on their L1 word-order cue (Isabelli, 2021). In her dissertation, Copeland (2022) found that advanced L1 English–L2 Spanish learners were just as accurate as Spanish native speakers at choosing the picture that correctly depicted a simple Spanish sentence. Proficiency was related to accuracy and RTs in the object–verb–subject (OVS) sentences. Higher proficiency late bilinguals were more accurate and faster in response to the OVS word order due to more exposure and experience in the L2 (Copeland, 2022). Thus, proficiency played a part in their L2 processing as well as L2 semantic knowledge.

Sagarra (2021) evaluated the differences between monolingual Spanish participants, low- and high-proficiency Romanian L1 learners of Spanish, and low- and high-proficiency English L1 learners of Spanish. In this eye-tracking study, the participants read two sentences, some with SV-agreement violations, and matched the sentences with a picture that best corresponded to the sentence. Sagarra (2021) found differences based on the participant's L1 and proficiency. Lower proficiency English participants relied less on determiners and more on the explicit subject than low-proficiency Romanian participants. Additionally, intermediate and advanced Romanian participants had longer verb dwell times than intermediate and advanced English participants. Advanced Romanian participants' eye behaviors on sentences with SV-agreement violations were more like those of native Spanish speakers than those of intermediate Romanian participants.

This highlights that typological similarities—both Spanish and Romanian being null-subject languages with rich verb agreement—aid L2 processing. When looking at differences in proficiency, Sagarra (2021) found that the lower the proficiency of the English participant, the more likely the participant was to rely on the noun to resolve verb-agreement violations. Sagarra (2021) concluded that the extent of acquisition of L2 cues largely depends on L2 proficiency, as late-age-of-acquisition participants were still able to acquire native-like cue processing with increased proficiency. In a summary of research on cue prediction, Karaca et al. (2021) found that language proficiency and, especially in the case of late bilinguals, L1/L2 similarities predicted successful predictions of cues. Because L1 proficiency is also associated with better L1 cue prediction, they theorized, but had not tested, that bilingual children would be able to predict cues more reliably if children had rich language input and exposure (Karaca et al., 2021). Accordingly, while there is research that links increased proficiency to more native-like cue processing in late L2 learners, there is a lack of understanding about how cue processing and proficiency are related in child and adult heritage speakers.

The second relevant processing type for simple sentences is lexical-based interpretations, which are also a highly salient method of determining thematic roles. Children can integrate lexico-semantic information into their parsing of a sentence. For English-speaking children, Bates et al. (1984) found that they will exclusively rely on word order unless there is an obvious semantic pairing. Children are able to combine semantic cues and syntactic processing. Additionally, in a recent study, Mahowald et al. (2023) found that meaning constrains the interpretation of a sentence and can render other superordinate cues, such as word order and agreement, redundant. All languages have built-in processing redundancy for processing rare, surprising, or ambiguous sentence meanings and facilitating language learning (Mahowald et al., 2023). In their study, they found that even with a strict and highly reliable word-order cue, semantic plausibility overrides the interpretation of the sentence (Mahowald et al., 2023). This cuing override is especially prevalent in sentences with two inanimate (IA-IA) nouns in which there is less reversibility of thematic roles without competing with semantic plausibility (Mahowald et al., 2023). The preceding research suggests that while bilinguals have competing cues across their languages, these cues might easily be dismissed due to the lexical-based content of a sentence. Thus, while the study reported here investigates the competition of the strongest cues from Spanish and English in bilinguals, semantic plausibility plays an important role in sentence processing, especially when there is an IA-IA pairing as is the case in the present study.

Historically, linguistic models explaining lexical-based representations and syntactic representations of sentence processing usually follow two paths: syntactic and lexical processing models are developed separately or in combination, wherein the syntactic frame is the main source of information, while lexical representation is only processed secondarily (Ferreira, 2003). However, the good-enough processing model (Ferreira et al., 2002) presents a contrasting view of sentence processing. While both CM and good-enough processing assume that the first noun encountered is the subject/agent, a good-enough approach to

language processing suggests two paths to a mental representation: one including heuristics and semantic processing and the other composed of syntactic processing (Ferreira, 2003). When linguistic stimuli are encountered, the language processing system enters a state of disequilibrium from the unprocessed, uncertain stimuli (Karimi and Ferreira, 2016). The language processing system will use heuristics and semantics first to arrive at equilibrium with an output but will use the syntactic algorithm as well if necessary for accuracy (Karimi and Ferreira, 2016). The language processing system is incentivized to process stimuli accurately but as quickly as possible to return to equilibrium. Thus, good-enough representations can occur even in the best of environments and even more so in noisy or irregular circumstances.

Studies on good-enough processing have traditionally targeted a syntactic structure that is considered difficult to understand. In their studies conducted exclusively on native English speakers, Ferreira et al. (2002) found that to understand passive sentences, participants rely more on semantic heuristics than on syntactic frames. When the syntactic structure does not follow the typical agent-first order, such as active compared to passive structures, a reliance on lexico-semantic plausibility is activated. Good-enough processing also suggests that semantic or syntactic representations are not always complete (Ferreira, 2003; Karimi and Ferreira, 2016). Even in experimental contexts in which the researchers manipulate context to lead participants to come to certain conclusions about the meaning of a sample, sometimes participants will come up with unexpected meanings. Sometimes participants create interpretations of sentences that do not match with syntactic structures, perhaps based on their past experiences or imagination. Evidence of good-enough processing can be seen even more so in noisy settings outside of the laboratory. Yet, stimuli will still be processed and understood despite errors in syntax and word choice. Ferreira et al. (2002) concluded that during processing, if the representation is not complete and supported with syntactic and semantic evidence, a good-enough interpretation might occur.

In a study comparing Korean–English bilinguals to English monolinguals, Lim and Christianson (2013) had participants read subject relative clauses (SRCs) and object relative clauses with plausible and implausible events, indicating a correct paraphrase, and, in a second session for the bilingual participants, providing a Korean translation. When indicating a correct paraphrase of the implausible SRCs, the L2 learners were less accurate than monolinguals (Lim and Christianson, 2013). The authors concluded this was due to L2 learner's reliance on semantic processing over syntactic cues. Additionally, in the translation task, there is more evidence of good-enough processing. When translating implausible relative clauses (RCs), participants would sometimes have syntactically accurate but thematically reversed roles in their translation (Lim and Christianson, 2013). This suggests that the syntactic parse was completed but the semantically implausible representation resulted in participants writing more sensible translations than the original RCs. Thus, Korean–English bilinguals elicited good-enough representations—when confronted with difficult and semantically implausible constructions—and in two different types of tasks. Given the aforementioned research showing IA-IA pairs are less reversible due to semantic implausibility (Mahowald et al., 2023), it is possible for good-enough representations to occur in the present study.

Based on the reviewed literature, it is evident that bilingual sentence processing involves a complex interplay between different cues and strategies, language proficiency, cue hierarchies, and the integration of lexico-semantic information. The studies conducted on Spanish–English bilinguals, heritage speakers, and late bilinguals highlight the variation in cue utilization and sentence interpretation patterns compared to monolinguals. Building upon these findings, the present study aims to investigate the cues employed by bilingual Spanish–English speakers when agreement and word-order cues are in competition. Additionally, the study aims to explore the influence of proficiency on cue interpretation RTs during sentence reading.

## Research questions

To address these research objectives, the following research questions and hypotheses are investigated based on CM and UCM theory (Bates and MacWhinney, 1989; MacWhinney, 2022):

1. What cues do bilingual Spanish–English speakers use when agreement and word-order cues are in competition with each other? Do cues vary by language input (English or Spanish sentences)?
2. How does proficiency affect cue interpretation RTs during sentence reading?

## Hypotheses

A. In English and Spanish, subject–verb–object (SVO) sentences will elicit faster RTs.
B. OVS sentences in English will elicit slower RTs, but participants will still overwhelmingly choose the traditional SVO word-order subject.
C. Participants that score higher on measures related to proficiency (the Lexical Test for Advanced Learners of English [LexTALE], the Lexical Test for Advanced Learners of English–Spanish [LexTALE-ESP], the Bilingual Language Profile [BLP], and Flanker) will elicit faster RTs to the Spanish OVS sentences than those with lower proficiency.
D. Participants of higher proficiency will choose the verb-agreeing subject (second noun) significantly more than those of lower proficiency on the OVS Spanish sentences.

## Methods

### Stimuli

To investigate what cues Spanish–English bilinguals use in both languages, the two strongest cues in the target languages, word order and verb agreement (Reyes and Hernández, 2006), were manipulated in sentences constructed to control for weaker cues such as animacy. While English only marks the accusative via word order and pronouns, Spanish marks the accusative on animate nouns via the accusative/personal *a* (Brugè and Brugger, 1996); this is called DOM. DOM is used with animate, specific

TABLE 1 Stimuli sample.

|  | Example sentence |
|---|---|
| English SVO | The faucet drips water. |
| English OVS | Sunlight absorb the solar panels. |
| Spanish SVO | El tornado destruye el pueblo. |
| Spanish OVS | Los tractores fabrica la factoría. |

This table shows examples of the sentences in each of the four conditions. SVO, subject–verb–object; OVS, object–verb–subject.

TABLE 2 Participant demographics.

|  |  | n (32) | % |
|---|---|---|---|
| Gender |  |  |  |
|  | Male | 18 | 56.25 |
|  | Female | 13 | 40.63 |
|  | Non-binary | 1 | 3.13 |
| Education |  |  |  |
|  | Less than high school | 0 | 0 |
|  | HS Degree/GED | 7 | 21.88 |
|  | Some College | 6 | 18.75 |
|  | College Degree (BA, BS) | 14 | 43.75 |
|  | Some graduate school | 1 | 3.13 |
|  | Masters | 4 | 12.5 |
|  | PhD/MD/JD | 0 | 0 |
| Household Income |  |  |  |
|  | $0–$19,999 | 2 | 6.25 |
|  | $20,000–$49,999 | 10 | 31.25 |
|  | $50,000–$89,999 | 10 | 31.25 |
|  | $90,000–$129,999 | 6 | 18.75 |
|  | $130,000–$149,000 | 1 | 3.13 |
|  | $150,000+ | 2 | 6.25 |
|  | Prefer not to answer | 1 | 3.13 |

Participant self-reported demographics based on categorical and ordinal options.

references ("men" as in mankind vs. "men" as in those men over there; Brugè and Brugger, 1996). Bates and MacWhinney (1989) listed the Spanish accusative *a* as the strongest cue for Spanish before verb agreement. Research that compares DOM production in monolinguals and heritage speakers found that monolinguals hardly ever omit the DOM for an animate, specific reference, but adult heritage speakers had significant rates of omission, averaging about 20% (Montrul and Sánchez-Walker, 2013). This also was mediated by individual factors; some heritage speakers were omitting all the time, and some were using the marker all the time, largely due to language exposure and birth order (Montrul and Sánchez-Walker, 2013). While heritage speakers do not seem to use the DOM cue to indicate an animate, specific direct object as much as monolingual speakers, it is still a salient cue. To limit the ability of participants to use a separate lexical cue to signal the direct object, the stimuli consisted of only inanimate nouns. This choice contrasts with the studies conducted by Copeland (2022) and Reyes and Hernández (2006) that similarly used simple sentences but animate nouns and chose to leave off DOM even when it would be ungrammatical.

Testing various word orders, despite ungrammaticality, is a common paradigm in the literature to see what cues are in use when word order is varied. While all six word orders were used in Reyes and Hernández (2006) study, this study will only focus on OVS and SVO word orders as most other word orders are not viable in either language. Nava (2007), Isabelli (2021), and Copeland (2022) all used a (O)VS word order in their studies on Spanish. Although an OVS order is valid in Spanish, it usually occurs with preverbal object pronouns (Nava, 2007; Isabelli, 2021). To keep the languages comparable, we will present SVO and OVS sentences both in Spanish and English, despite OVS not being a viable word order in English. Therefore, the stimuli, as sampled in Table 1 and presented in detail in Appendix A, are composed of present-tense SVO and OVS sentences in English and Spanish. The 80 unique sentences were split between the languages with 40 sentences in English and 40 sentences in Spanish. Among these, 20 sentences followed the SVO word order, while the other 20 sentences followed the OVS word order in both languages.

## Procedure

Thirty-five participants were recruited for the study via Prolific.com, an online platform for participant recruitment. Three participants who had acquired English later, that is, in middle school or after the onset of puberty (in this case, 13–14 years old)

were excluded from the data analysis. In Tables 2, 3, it can be seen that we had a diverse sample, with more males than females. Participants were provided with monetary compensation for their participation. Only computers, not tablets or mobile devices, were allowed for this study. We set the participation criteria in Prolific as participants who had acquired Spanish as a first language and were fluent in both English and Spanish. Participants who were eligible, and chose to participate, were directed to the Gorilla Experiment Builder platform, where the experiment was conducted. Before to participating, potential participants were screened and presented with the following question in both English and Spanish: "Are you a Spanish–English bilingual, whose first language was Spanish?" Only those who met this criterion were eligible to proceed by saying "Yes," and those who responded "No" were redirected back to Prolific. Table 3 summarizes the participants' language background. Study participants acquired Spanish at the same time or earlier than they acquired English. Additionally, the majority of participants received more education in English than in Spanish, suggesting that they were in an English-dominant environment. Upon meeting the eligibility criteria, participants were presented with a consent form in English, which they were required to read and then provide their informed consent before proceeding with the study. If they declined consent, they were redirected back to Prolific.

TABLE 3 Participant age and language acquisition age and education.

| Age | | |
|---|---|---|
| | Mean (SD) | 33 (9.29) |
| | Range | 19–57 |
| English Acq. Age | | |
| | Mean (SD) | 3.5 (2.42) |
| | Range | 0–9 |
| Spanish Acq. Age | | |
| | Mean (SD) | 1.03 (0.82) |
| | Range | 0–3 |
| English Ed. Years | | |
| | Mean (SD) | 11.94 (4.44) |
| | Range | 0–18 |
| Spanish Ed. Years | | |
| | Mean (SD) | 5.47(4.82) |
| | Range | 0–18 |

Descriptive statistics of participant's age at experimentation, age of language acquisition, and years of language education. Acq., acquisition; ed., education.

The following instructions were then provided to participants before their training task and both sections of the main task. "For the following task, read the sentences and decide the subject of the sentence. Once you have read the sentence press the SPACE bar. First, you will see a fixation cross, after a few seconds it will automatically advance to the next screen. On this next screen, please click noun that you think was the subject of the sentence."[1]. Below the English instructions, the same instructions were given in Spanish. Prior to the main task, participants completed a practice session that consisted of two practice items per each of the four treatments. Participants were randomly assigned into two groups: one group received the English block first, while the other group received the Spanish block first. Following the completion of their first block, participants then received the block that they had not received initially, resulting in a total of 80 sentences per participant. The order of completing these tests was randomized to control for any potential order effects. The task itself involved determining the subject of each sentence. The stimuli were presented on a computer screen in the default Gorilla font, Inter var ExtraLight 200 size 50 for the sentences and size 60 for the choice selection as shown in Figure 1. The participants were presented with a fixation cross (1,100 ms), the sentence, a fixation cross (1,100 ms), and then the noun choices. The participants could progress from the sentence-reading screen via a press of the spacebar, and then, after a quick fixation screen (1,100 ms), they were presented with the noun choices. The sentences were randomized within each language. Following the

---

1 Minor grammatical errors were found in the English instructions of the main task after administration of the study. However, we are confident that this did not overly affect the participants as these errors were not found in the Spanish instructions, which appeared right below the English instructions, and they had eight practice trials before the main task.

sentence processing task, the participants were offered a break screen. When they were done with their break, they could continue the rest of the study, which included three measures related to language proficiency.

## Measures

### Language proficiency

The LexTALE (Lemhöfer and Broersma, 2012) and the LexTALE-ESP (Izura et al., 2014) were used as measures of language proficiency. Both are measures of receptive vocabulary and, as such, indirect measures of language proficiency. The LexTALE is composed of 60 items in which the participant must indicate whether a word is an actual word or a non-word. It takes approximately 4 min to complete (Lemhöfer and Broersma, 2012). Likewise, the LexTALE-ESP takes approximately 5 min and has 90 test items. Both tests vary the frequency of the words tested and created non-words based on the phonotactics of the respective language. Both tests have been shown to be reliable measures of receptive vocabulary size, which is positively correlated with proficiency (Lemhöfer and Broersma, 2012; Izura et al., 2014). The order of the proficiency assessments was randomized to control for any potential order effects.

### BLP

Finally, participants were asked to give demographic data on education, income, age, and occupation and complete the BLP (Birdsong et al., 2012). The BLP is a measure of language dominance that is related to language proficiency and usage (Bonvin et al., 2021). The BLP takes approximately 10 min to complete and gathers information about language history, use, proficiency, and attitudes for both languages (Birdsong et al., 2012). In the section on language history, participants self-report when they started learning the language, how many years of education received education in the language, and how long they have been in a family or work environment with the language on a scale of 0–20+ years. The next section asks participants to give a percentage of time that they use each language in five different contexts. Then participants are asked to indicate on a Likert scale their language abilities in speaking, listening, reading, and writing. Finally, in a set of four questions, participants indicate how much they identify with statements about the language, culture, and being a native speaker or not. These sections all have individual scoring metrics and weights.

### Flanker task

Language processing and prediction is related to domain-general cognitive skills, such as working memory, attention, and the like (Karaca et al., 2021). Declines in cognitive skills associated with aging also lead to declines in language processing and prediction (Karaca et al., 2021). Much work has been done in the field of psycholinguistics to link bilingualism with improvements in cognitive skills, also called the "bilingual advantage" debate.
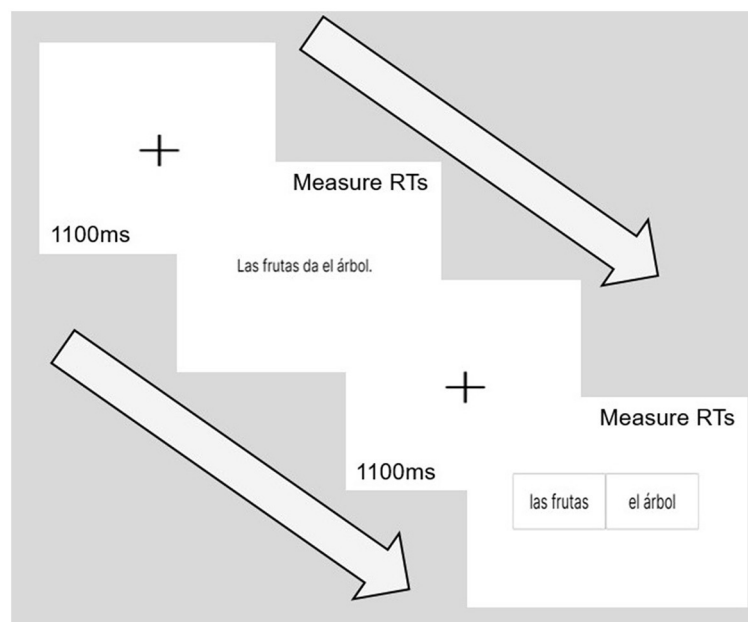
**FIGURE 1**
Timeline of main task. Visualization of the experimental setup of the main task. RT, reaction time.

Meta-analyses of over 100 studies have found significant differences in bilingual performance on cognitive tasks targeting executive functions when compared to monolinguals (Grundy, 2020; Ware et al., 2020). It is widely reported that bilinguals have increased inhibitory control due to consistently suppressing the activation of the unnecessary language(s) (Luk et al., 2011). Additionally, research suggests that higher proficiency and usage of both languages is associated with higher inhibitory control, as demonstrated by higher accuracy and faster response time on incongruent trials on the Flanker task (Luk et al., 2011). Critics against the bilingual advantage in executive functions cite a lack of control of matched socio-economic status (SES) participants, however, in a study on low-SES bilinguals (Thomas-Sunesson et al., 2018), a difference in inhibitory control was still found between more balanced participants and more asymmetrical participants. Specifically, Spanish–English heritage children who were more balanced in their language proficiency received more benefit to their inhibitory control than participants who were more proficient in one language over the other (Thomas-Sunesson et al., 2018). It is worth noting, however, that Ware et al. (2020) meta-analysis did not find a significant effect of the Flanker task, but such an effect was found for a similar task, the Attentional Network Task. In this study, the Flanker task was used as a measure of inhibitory control and an indirect measure of proficiency. Participants were required to complete a Flanker task of 150 trials, which consisted of 50 iterations per neutral (- - > - -), congruent (> > > > >), and non-congruent (< < > < <) items. Before the Flanker task, participants received instructions and completed a nine-item practice session. The participants were instructed to press "F" if the center arrow was pointing to the left or "J" if the center arrow was pointing to the right. After a set of 50 trials, participants were given a break and then given three practice trials that were not scored.

## Results

## Statistical analyses

All analyses were conducted in R (v4.2.3; R Core Team, 2023). Linear mixed-effects (LME) regression and generalized linear mixed-effects models (GLMMs) were fitted using the lme4 and lmerTest packages (Bates et al., 2015; Kuznetsova et al., 2017). The statistical results were written with the help of the report package (Makowski et al., 2023). RTs were clustered and centered by language and word order. The four clusters were averaged and then the cluster mean was subtracted from the value. A constant value was then added to allow for a log transformation to be performed on the cluster-centered RTs. The LexTALE and LexTALE-ESP were scored via the averaged percentage correct method indicated by Lemhöfer and Broersma (2012) and then were $z$-scored to have comparable scales in data analysis. The BLP was scored according to traditional scoring methods (Birdsong et al., 2012) and then standardized. The Flanker task was scored by taking the RTs of the correct congruent trials from the correct incongruent trials for an interference RT. Then the interference RT was scaled to have a mean of 0 and a standard deviation of 1. In the following models, LME $p$-values were computed using a Wald $t$-distribution approximation and GLMM $p$-values were computed using a Wald $z$-distribution approximation. The models can be compared side by side in Table 4. Figures 2, 3 and Table 4 were created using the following packages: ggpubr, ggplot2, and tidyverse (Wickham, 2016; Wickham et al., 2019; Kassambara, 2023).

## Subject choice selection models

To address the first research question regarding which cues participants were attending to, we fitted two GLMM or logistic

TABLE 4  Model comparison for subject choice and reaction time.

| Predictors | SP-Subject Choice | | | EN-Subject Choice | | | SP-Sentence RT | | | EN-Sentence RT | | | SP-Choice RT | | | EN-Choice RT | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Odds ratios | CI | p | Odds ratios | CI | p | Estimates | CI | p | Estimates | CI | p | Estimates | CI | p | Estimates | CI | p |
| (Intercept) | 0.12 | [0.07, 0.20] | **<0.001** | 0.05 | [0.03, 0.09] | **<0.001** | 8.15 | [7.98, 8.32] | **<0.001** | 7.94 | [7.81, 8.08] | **<0.001** | 7.35 | [7.23, 7.46] | **<0.001** | 7.38 | [7.30, 7.47] | **<0.001** |
| OVS | 13.33 | [9.67, 18.38] | **<0.001** | 36.62 | [24.66, 54.39] | **<0.001** | −0.17 | [−0.30, −0.05] | **0.005** | −0.01 | [−0.11, 0.09] | 0.858 | −0.65 | [−0.72, −0.58] | **<0.001** | −0.06 | [−0.11, −0.01] | **0.017** |
| LexTALE | 1.07 | [0.65, 1.77] | 0.786 | 0.66 | [0.42, 1.05] | 0.078 | −0.16 | [−0.33, 0.01] | 0.072 | −0.17 | [−0.31, −0.03] | **0.017** | −0.19 | [−0.30, −0.08] | **0.001** | −0.12 | [−0.20, −0.03] | **0.006** |
| LexTALE-ESP | 1.13 | [0.67, 1.90] | 0.658 | 1.45 | [0.90, 2.34] | 0.128 | 0.05 | [−0.13, 0.23] | 0.560 | 0.09 | [−0.06, 0.23] | 0.229 | −0.06 | [−0.18, 0.06] | 0.301 | 0.01 | [−0.08, 0.09] | 0.847 |
| BLP | 0.74 | [0.44, 1.25] | 0.262 | 0.92 | [0.57, 1.49] | 0.736 | 0.25 | 0.07 – 0.43 | **0.007** | 0.12 | [−0.02, 0.27] | 0.099 | 0.11 | [−0.01, 0.23] | 0.062 | 0.11 | [0.02, 0.19] | **0.015** |
| Flanker | 0.84 | [0.51, 1.40] | 0.507 | 0.99 | [0.62, 1.57] | 0.960 | −0.06 | [−0.23, 0.12] | 0.534 | 0.02 | [−0.12, 0.17] | 0.736 | 0.01 | [−0.10, 0.13] | 0.833 | −0.04 | [−0.12, 0.05] | 0.368 |
| **Random effects** | | | | | | | | | | | | | | | | | | |
| $\sigma^2$ | 3.29 | | | 3.29 | | | 0.59 | | | 0.49 | | | 0.45 | | | 0.20 | | |
| $\tau_{00}$ | 1.77 Participant | | | 1.42 Participant | | | 0.21 Participant | | | 0.13 Participant | | | 0.09 Participant | | | 0.05 Participant | | |
| $\tau_{11}$ | | | | | | | 0.07 Participant.WordOrder | | | 0.03 Participant.WordOrder | | | | | | | | |
| $\rho_{01}$ | | | | | | | 0.55 Participant | | | 0.64 Participant | | | | | | | | |
| ICC | 0.35 | | | 0.30 | | | 0.34 | | | 0.27 | | | 0.16 | | | 0.20 | | |
| N | 32 Participant | | | 32 Participant | | | 32 Participant | | | 32 Participant | | | 32 Participant | | | 32 Participant | | |
| Marginal $R^2$/Conditional $R^2$ | 0.267/0.523 | | | 0.433/0.604 | | | 0.088/0.400 | | | 0.059/0.313 | | | 0.227/0.352 | | | 0.097/0.275 | | |
| AICc | 1,239.669 | | | 1,062.428 | | | 3,097.127 | | | 2,861.800 | | | 2,700.100 | | | 1,656.903 | | |

The "SP-Subject Choice" and "EN-Subject Choice" models are generalized linear mixed models (GLMM), whereas the remaining models are linear mixed effects models (LME). The bold values are statistically significant results. SP, Spanish; EN, English; RT, reaction time; OVS, object–verb–subject; LexTALE, Lexical Test for Advanced Learners of English; LexTALE-SP, Lexical Test for Advanced Learners of English–Spanish; BLP, Bilingual Language Profile; ICC, Intraclass correlation; AICc, Akaike information criterion-corrected.

mixed models, one for each language (estimated using ML and BOBYQA optimizer), to predict subject choice with word order, LexTALE score, LexTALE-ESP score, BLP score, and Flanker Interference RT with the following R code: subject choice ∼ word order + LexTALE score + LexTALE-ESP score + BLP score + Flanker Interference RT. The models included ParticipantID as a random effect (code: ∼1 | ParticipantID)[2].

### Spanish

The model's total explanatory power is substantial (conditional $R^2 = 0.52$), and the part related to the fixed effects alone (marginal $R^2$) is 0.27. The model's intercept, corresponding to word order = SVO, LexTALE score = 0, LexTALE-ESP score = 0, BLP score = 0, and Flanker Interference RT = 0, is at −2.13 (95% CI [−2.66, −1.60], $p < 0.001$). The effect of word order OVS is statistically significant and positive (beta = 2.59, 95% CI [2.27, 2.91], $p < 0.001$; Std. beta = 1.30, 95% CI [1.14, 1.46]). In the OVS condition, there is a 2.59 log-odds likelihood or 93.02% probability of choosing the second noun or the verb-agreeing noun. The effect of LexTALE score is statistically non-significant and positive (beta = 0.07, 95% CI [−0.43, 0.57], $p = 0.786$; Std. beta = 0.07, 95% CI [−0.43, 0.57]). The effect of LexTALE-ESP score is statistically non-significant and positive (beta = 0.12, 95% CI [−0.40, 0.64], $p = 0.658$; Std. beta = 0.12, 95% CI [−0.40, 0.64]). The effect of BLP score is statistically non-significant and negative (beta = −0.30, 95% CI [−0.83, 0.23], $p = 0.262$; Std. beta = −0.30, 95% CI [−0.83, 0.23]). LexTALE, LexTALE-ESP, and BLP scores are not significant predictors in choosing the subject of the sentence. The effect of Flanker Interference RT is statistically non-significant and negative (beta = −0.17, 95% CI [−0.68, 0.33], $p = 0.507$; Std. beta = −0.17, 95% CI [−0.68, 0.33]).

### English

The model's total explanatory power is substantial (conditional $R^2 = 0.60$), and the part related to the fixed effects alone (marginal $R^2$) is 0.43. The model's intercept, corresponding to word order = SVO, LexTALE score = 0, LexTALE-ESP score = 0, BLP score = 0, and Flanker Interference RT = 0, is at −2.97 (95% CI [−3.50, −2.43], $p < 0.001$). The effect of word order OVS is statistically significant and positive (beta = 3.60, 95% CI [3.21, 4.00], $p < 0.001$; Std. beta = 1.80, 95% CI [1.60, 2.00]). In the OVS condition, there is a 3.60 log-odds likelihood or 97.3% probability of choosing the second noun. The effect of LexTALE score is statistically non-significant and negative (beta = −0.41, 95% CI [−0.87, 0.05], $p = 0.078$; Std. beta = −0.41, 95% CI [−0.87, 0.05]). The effect of LexTALE-ESP score is statistically non-significant and positive (beta = 0.37, 95% CI [−0.11, 0.85], $p = 0.128$; Std. beta = 0.37, 95% CI [−0.11, 0.85]). The effect of BLP score is statistically non-significant and negative (beta = −0.08, 95% CI [−0.56, 0.40], $p = 0.736$; Std. beta = −0.08, 95% CI [−0.56, 0.40]). The effect of Flanker Interference RT is statistically non-significant and negative (beta = −0.01, 95% CI [−0.47, 0.45], $p = 0.960$; Std. beta = −0.01, 95% CI [−0.47, 0.45]). Like with Spanish choice selection, LexTALE, LexTALE-ESP, BLP scores, and Flanker Interference RT are not significant predictors.

## RT models

To answer our second research question about whether proficiency affected RTs, we fitted four linear mixed models, one for each language by both sentence or choice reading (estimated using REML and nloptwrap optimizer), to predict RT with word order, LexTALE score, LexTALE-ESP score, BLP score, and Flanker Interference RT (code: RT ∼ word order + LexTALE score + LexTALE-ESP score + BLP score + Flanker Interference RT). The sentence RT models included word order as random effects (code: ∼1 + word order | ParticipantID)[3]. This model, which allowed for the possible variation between word order and participant, fit the data significantly better than the model that did not contain a random slope. The subject choice selection RT models included ParticipantID as a random effect (code: ∼1 | ParticipantID)[4].

### Sentence RT models
#### Spanish

The overall model explained 40% of the variation in participant responses (conditional $R^2 = 0.40$) and the part related to the fixed effects alone (marginal $R^2$) is of 0.09. The model's intercept, corresponding to word order = SVO, LexTALE score = 0, LexTALE-ESP score = 0, BLP score = 0, and Flanker Interference RT = 0, is at 8.15 (95% CI [7.98, 8.32], $t_{(1,270)} = 94.50$, $p < 0.001$). Within this model, the effect of word order OVS is statistically significant and negative (beta = −0.17, 95% CI [−0.30, −0.05], $t_{(1,270)} = −2.79$, $p = 0.005$; Std. beta = −0.09, 95% CI [−0.15, −0.03]). Thus, the OVS word order predicts faster RTs in reading Spanish sentences. As seen in Figure 2, the Spanish OVS condition had a much wider range of RTs than any of the other conditions. The effect of LexTALE score is nearly statistically significant and negative (beta = −0.16, 95% CI [−0.33, 0.01], $t_{(1,270)} = −1.80$, $p = 0.072$; Std. beta = −0.16, 95% CI [−0.34, 0.01]). The effect of LexTALE-ESP score is statistically non-significant and positive (beta = 0.05, 95% CI [−0.13, 0.23], $t_{(1,270)} = 0.58$, $p = 0.560$; Std. beta = 0.06, 95% CI [−0.13, 0.24]). The effect of BLP score is statistically significant and positive (beta = 0.25, 95% CI [0.07, 0.43], $t_{(1,270)} = 2.72$, $p = 0.007$; Std. beta = 0.26, 95% CI [0.07, 0.44]). Participants who are more English-dominant are slower at reading Spanish sentences. The effect of Flanker Interference RT is statistically non-significant and negative (beta = −0.06, 95% CI [−0.23, 0.12], $t_{(1,270)} = −0.62$, $p = 0.534$; Std. beta = −0.06, 95% CI [−0.24, 0.12]).

### English

The model's total explanatory power is substantial (conditional $R^2 = 0.31$), and the part related to the fixed effects alone (marginal $R^2$) is 0.06. The model's intercept, corresponding to word order = SVO, LexTALE score = 0, LexTALE-ESP score = 0, BLP score = 0, and Flanker Interference RT = 0, is at 7.94 (95% CI [7.81, 8.08], $t_{(1,270)} = 115.08$, $p < 0.001$). The effect of word order OVS is statistically non-significant and negative (beta = −8.93e-03, 95%

---

2   $\text{logit}(\pi_i) = \beta_0 + \beta_1 \times \text{WordOrderOVS}_i + \beta_2 \times \text{LexTALE}_i + \beta_3 \times \text{LexTALE-ESP}_i + \beta_4 \times \text{BLP}_i + \beta_5 \times \text{Flanker}_i + u_{j[i]}$.

3   $\text{Sentence RT}_i = \beta_0 + \beta_1 \times \text{WordOrderOVS}_i + \beta_2 \times \text{LexTALE}_i + \beta_3 \times \text{LexTALE-ESP}_i + \beta_4 \times \text{BLP}_i + \beta_5 \times \text{Flanker}_i + (u_{0j} + u_{1j} \times \text{WordOrder}_i) + \epsilon_i$.

4   $\text{Subject Choice RT}_i = \beta_0 + \beta_1 \times \text{WordOrderOVS}_i + \beta_2 \times \text{LexTALE}_i + \beta_3 \times \text{LexTALE-ESP}_i + \beta_4 \times \text{BLP}_i + \beta_5 \times \text{Flanker}_i + u_{j[i]} + \epsilon_i$.
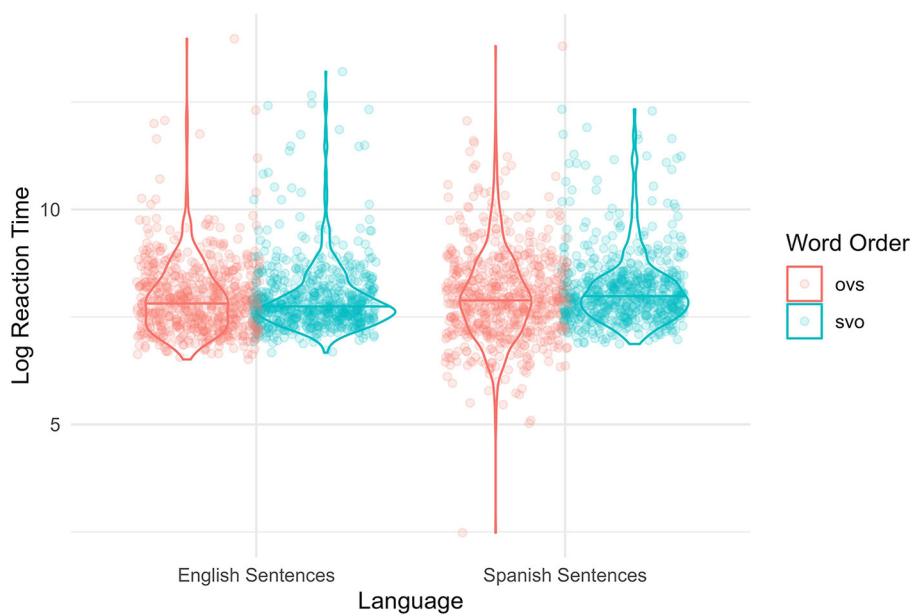
**FIGURE 2**
Distribution of sentence reaction time. Participants' log-transformed, cluster-centered reaction times while reading the stimuli sentences across the four clusters. OVS, object−verb−subject; SVO, subject−verb−object.

CI $[-0.11, 0.09]$, $t_{(1,270)} = -0.18$, $p = 0.858$; Std. beta $= -5.35e\text{-}03$, 95% CI $[-0.06, 0.05]$). Unlike with the Spanish condition, the OVS word order is not significantly different from the SVO word order. Examining Figure 2, the similarities in the distribution of the RTs to the English sentences between the two word orders can be seen. The effect of LexTALE score is statistically significant and negative (beta $= -0.17$, 95% CI $[-0.31, -0.03]$, $t_{(1,270)} = -2.38$, $p = 0.017$; Std. beta $= -0.20$, 95% CI $[-0.37, -0.04]$). The effect of LexTALE-ESP score is statistically non-significant and positive (beta $= 0.09$, 95% CI $[-0.06, 0.23]$, $t_{(1,270)} = 1.20$, $p = 0.229$; Std. beta $= 0.11$, 95% CI $[-0.07, 0.28]$). Thus, for the English sentences, English vocabulary size or proficiency predicts faster RTs, and Spanish vocabulary size or proficiency is not a significant predictor. The effect of BLP score is non-significant and positive (beta $= 0.12$, 95% CI $[-0.02, 0.27]$, $t_{(1,270)} = 1.65$, $p = 0.099$; Std. beta $= 0.15$, 95% CI $[-0.03, 0.32]$). This pattern is similar to the Spanish model but not quite significant, BLP slows the RTs, but to a lesser degree. As the BLP increases by one level, RT also increases. Finally, the effect of Flanker Interference RT is statistically non-significant and positive (beta $= 0.02$, 95% CI $[-0.12, 0.17]$, $t_{(1,270)} = 0.34$, $p = 0.736$; Std. beta $= 0.03$, 95% CI $[-0.14, 0.20]$).

## Subject choice RT models

### Spanish

The model's total explanatory power is substantial (conditional $R^2 = 0.35$) and the part related to the fixed effects alone (marginal $R^2$) is 0.23. The model's intercept, corresponding to word order $=$ SVO, LexTALE score $= 0$, LexTALE-ESP score $= 0$, BLP score $= 0$, and Flanker Interference RT $= 0$, is at 7 7.35 (95% CI $[7.23, 7.46]$, $t_{[1,272]} = 126.33$, $p < 0.001$). The effect of word

order OVS is statistically significant and negative, predicting faster RTs in choosing a subject (beta $= -0.65$, 95% CI $[-0.72, -0.58]$, $t_{[1,272]} = -17.39$, $p < 0.001$; Std. beta $= -0.40$, 95% CI $[-0.44, -0.35]$). In Figure 3, log RTs appear on the $x$-axis. It is evident the OVS distribution of RTs is skewed slightly to the left compared to the SVO distribution. The effect of LexTALE score is statistically significant and negative (beta $= -0.19$, 95% CI $[-0.30, -0.08]$, $t_{[1,272]} = -3.28$, $p = 0.001$; Std. beta $= -0.23$, 95% CI $[-0.36, -0.09]$). The effect of LexTALE-ESP score is statistically non-significant and negative (beta $= -0.06$, 95% CI $[-0.18, 0.06]$, $t_{[1,272]} = -1.03$, $p = 0.301$; Std. beta $= -0.08$, 95% CI $[-0.22, 0.07]$). The sentence-reading RTs, when choosing a subject in Spanish, increased LexTALE score predicts faster RTs, but LexTALE-ESP scores do not. The effect of BLP score is nearly statistically significant and positive (beta $= 0.11$, 95% CI $[-5.75e\text{-}03, 0.23]$, $t_{[1,272]} = 1.87$, $p = 0.062$; Std. beta $= 0.14$, 95% CI $[-7.02e\text{-}03, 0.28]$). The effect of Flanker Interference RT is statistically non-significant and positive (beta $= 0.01$, 95% CI $[-0.10, 0.13]$, $t_{[1,272]} = 0.21$, $p = 0.833$; Std. beta $= 0.01$, 95% CI $[-0.12, 0.15]$).

### English

The model's total explanatory power is substantial (conditional $R^2 = 0.28$) and the part related to the fixed effects alone (marginal $R^2$) is 0.10. The model's intercept, corresponding to word order $=$ SVO, LexTALE score $= 0$, LexTALE-ESP score $= 0$, BLP score $= 0$, and Flanker Interference RT $= 0$, is at 7.38 (95% CI $[7.30, 7.47]$, $t_{[1,272]} = 173.75$, $p < 0.001$). The effect of word order OVS is statistically significant and predicted faster RTs (beta $= -0.06$, 95% CI $[-0.11, -0.01]$, $t_{[1,272]} = -2.40$, $p = 0.017$; Std. beta $= -0.06$, 95% CI $[-0.11, -0.01]$). This relationship is seen in Figure 4 where the OVS RTs distribution is shifted to the left in comparison to the
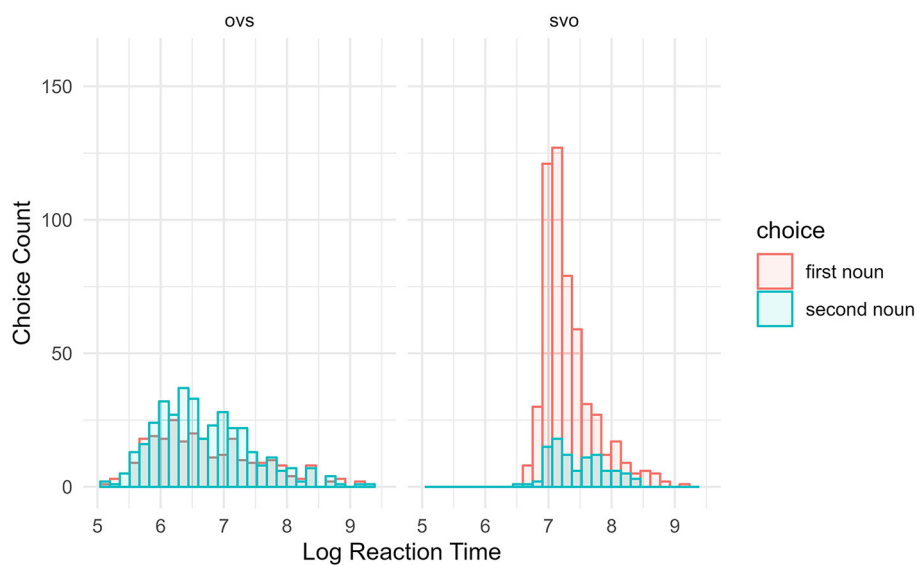
FIGURE 3
Spanish subject choice reaction time by word order. Participants' log-transformed, cluster-centered reaction times while making a subject selection on the Spanish stimuli.
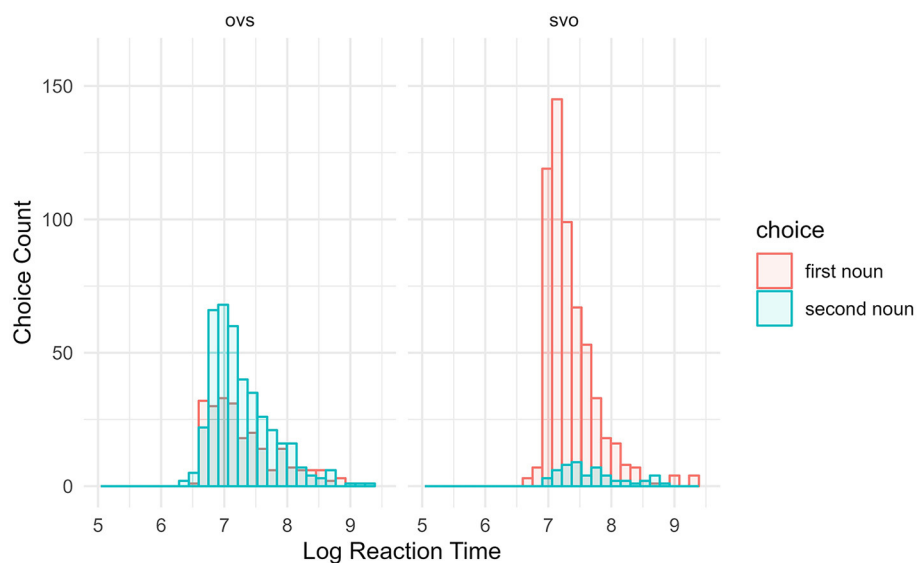


FIGURE 4
English subject choice reaction time by word order. Participants' log-transformed, cluster-centered reaction times while making a subject selection on the English stimuli.

SVO RTs. The effect of LexTALE score is statistically significant and negative (beta = −0.12, 95% CI [−0.20, −0.03], $t_{[1,272]} = -2.75$, $p = 0.006$; Std. beta = −0.23, 95% CI [−0.39, −0.06]). Greater English vocabulary size leads to faster RTs in English sentences. The effect of LexTALE-ESP score is statistically non-significant and positive (beta = 8.48e-03, 95% CI [−0.08, 0.09], $p = 0.19$, $p = 0.847$; Std. beta = 0.02, 95% CI [−0.15, 0.19]). Here we see a similar pattern to the Spanish choice RT results: LexTALE scores

predict faster RTs in choosing a subject in English sentences, but LexTALE-ESP scores do not. The effect of BLP score is statistically significant and positive (beta = 0.11, 95% CI [0.02, 0.19], $t_{[1,272]} = 2.43$, $p = 0.015$; Std. beta = 0.21, 95% CI [0.04, 0.38]). Increases toward English dominance lead to slower RTs in English sentences. The effect of Flanker Interference RT is statistically non-significant and negative (beta = −0.04, 95% CI [−0.12, 0.05], $t_{[1,272]} = -0.90$, $p = 0.368$; Std. beta = −0.08, 95% CI [−0.24, 0.09]).

## Discussion

To understand how Spanish–English heritage speakers have integrated two of the strongest cues in their languages, we analyzed RT for sentence reading and choice selection and looked at the difference in the binary subject choice in both English and Spanish. To answer our first research question about which cue is being used when the two cues are in competition, we ran two GLMM models to look at which noun the participants would pick in each language. Generally, we predicted that if participants were utilizing the word-order cue, they would choose the first noun significantly more than the second noun in the OVS conditions. In contrast, if they were using the agreement cue, they would pick the second noun more in the OVS Spanish condition. However, these predictions—along with their corresponding proficiency predictions—were not observed.

For the English OVS sentences, participants selected the second noun the majority of the time. Our prediction assumed that even with an amalgamation of cues from both languages, participants in the English condition would still default to choosing the first noun. However, participants appeared to be more evenly split in their processing of the OVS word order than their processing of the SVO word order. In the English OVS condition, the distinction between third-person singular and third-person plural verb agreement is only that of one character (-s). Therefore, participants could have mentally corrected for, skipped, or thought that the -s was a mistake. This could have easily happened in sentences such as "Colors changes the traffic light." A little less than half of the OVS English sentences had a third-person plural as the subject to balance the -s deletion; however, mental correction could still be happening. If participants are mentally correcting the OVS sentences to, for example, "Colors change the traffic light," participants would need to be suspending semantic plausibility but would be relying on English's strong word-order cue. This could account for a small percentage of participants choosing the first noun as the subject in the OVS condition, with this sentence being the most likely culprit as *colors* can *change*. For this first example, participants would not know they have misprocessed the OVS sentence until they read the second noun. If most of the OVS sentences were similar, perhaps the OVS condition would have been slower. This mistake is less likely to happen for OVS sentences such as "The pants smooths the iron," because *pants* do not *smooth*. Sentences like these, where the participant knows that the first noun is not the agent of the verb immediately upon reading the verb, are much more frequent in the stimuli. Thus, participants may be using other cues than syntactic cues to complete the task.

Indeed, based on the UCM (MacWhinney, 2013, 2022) and the findings from Mahowald et al. (2023), there could be two interpretations of this result. When recruiting for this study, we specifically looked for Spanish heritage speakers, or speakers that had learned Spanish first in the home. Thus, for these participants, Spanish was their first language, although, likely due to external factors and influences, the majority of participants are more proficient and dominant in English (LexTALE $M = 89.65\%$, BLP-English $M = 185.7/218$) than Spanish (LexTALE-ESP $M = 67.76\%$, BLP-Spanish $M = 151.65/218$). Therefore, what we see in this context is likely forward transfer, not amalgamation. Participants

may be using the L1 Spanish agreement cues on L2 English OVS sentences. However, the results from the Spanish GLMM do not correspond well to this explanation. In the Spanish sentence condition, there was a 93% (compared to the 97.3%) probability of participants selecting the second noun in the OVS word order. For forward transfer to be the most probable explanation, we would expect to see at least the same or higher likelihood of choosing the agreeing noun in Spanish over English. Thus, the agreement cue, which is stronger in Spanish than in English (Reyes and Hernández, 2006), does not appear to provide a strong explanation for these results. For the same reason, we do not favor a differentiation explanation. If there had been a differentiation of cues, we would have expected that in the English OVS condition, our participants would have chosen the first noun, corresponding to English's strong word-order cue.

Mahowald et al. (2023) suggest that IA-IA sentences, like those used in this study, are less likely to be reversible. Despite word-order shifts, IA-IA sentences are limited in how thematic roles can be assigned due to semantic relationships, such as some IA nouns being more agentive. Thus, participants who have more knowledge of English semantics are more likely to pick the correct subject according to the semantics of the words in the sentence than they can in Spanish, due to lower Spanish proficiency. This accounts for the higher probability of choosing the second noun, or the agreeing subject, in English (at 97.3%), in comparison to Spanish (at 93%). While this seems to be a more sensible explanation of the data, further studies are needed to test this explanation fully. We propose future studies to test this: first, testing IA-IA sentences in different word orders or using words in lemma forms and not in sentence structure as in Mahowald et al. (2023) to lessen the effect of syntax or, second, including more proficiency measures specifically targeting semantic knowledge in both languages, since the LexTALE-ESP might not be a sufficiently precise enough tool for our purposes.

Nevertheless, this lexico-semantic-based conclusion is also supported by the good-enough processing model (Ferreira et al., 2002; Ferreira, 2003; Karimi and Ferreira, 2016). While typical readings of sentences may trigger the first-noun-as-agent heuristic, the first nouns of these OVS sentences are not typical agents or subjects in sentences. In this study, participants could have begun a semantic reading of the sentence, merely reading for meaning and plausibility, and picked the appropriate noun from that analysis. However, if we accept this position, we would be assuming, based on the results of this study, that reading for a semantic representation is faster than reading for a syntactic representation. This is supported in electroencephalogram (EEG) studies that show lexical conflicts are triggered at the N400 before P600 syntactic conflicts (Frenzel et al., 2011). Frenzel et al. (2011) found that the N400 was specifically induced when there were violations in assigning the actor thematic role, but it also was not as large with inanimate nouns. Furthermore, work with IA-IA longer sentences and eye tracking to examine the reading patterns of Spanish–English bilinguals would address this issue more directly, or in EEG, to examine the N400 response to IA-IA actor violations.

From these GLMMs, we can also test our hypothesis on whether participants with higher Spanish proficiency would choose the verb-agreeing noun. While there was an overall increased

likelihood of choosing the second noun, or the agreeing noun, in the OVS word order, proficiency did not have a significant effect on the choice of nouns in either language. Because proficiency was not significant in these models, the hypothesis is not supported. As seen in studies on late bilinguals (Isabelli, 2021; Copeland, 2022), we would have expected to see at least proficiency, if not dominance, to play a role if subject choice were due to forward transfer. Neither variable significantly predicted word selection. However, the claim that results are due to the inherent nature of IA-IA sentences and possible good-enough processing cannot fully be supported without further studies.

To answer Research Question 2 regarding the impact of proficiency on RTs, we performed two sets of LMEs, with cluster-centered, log-transformed RTs. In the LMEs for sentence RTs, we found that the participants read the Spanish OVS sentences faster. Forward transfer of the Spanish agreement cue can explain these results as well as the prior studies showing IA-IA sentences are less reversible. Although with either of these, we expected the RTs to be similar across word-order conditions, not faster in the OVS condition in Spanish. Our expectation was the agreement would reduce RTs in the OVS condition to be on par with those in the SVO condition, considering the OVS is not a typical word order in Spanish. Previous research has found that, in comparison to Spanish monolinguals, Spanish–English bilinguals tend to prefer SV constructions to VS constructions (Nava, 2007), the forward-transfer explanation for the faster OVS RTs is not strongly supported. If IA-IA sentences being less reversible were the only explanation, we would expect slower RTs in the OVS condition, as participants reread the sentences upon realizing they had misprocessed. However, if we were seeing a differentiation of cues, we would expect that the OVS word order would be much slower in English as it is a non-canonical word order and participants would not know they had misprocessed the sentence until they had read the verb or the second noun. However, the word-order RTs in English were not significantly different, leading us to seek another explanation.

In these sentence RT LMEs, we also see that an increase in English proficiency leads to a decrease in reading time in English and a near-significant decrease in reading time in Spanish. While this result makes sense in the English condition, this pattern was not expected in the Spanish condition. Paired with the result that an increase in a BLP score, associated with English dominance, corresponds to a significant increase in RTs in the Spanish sentences, the picture here is quite complex. We consider two interpretations for this pattern of results. First, this pattern could be due to the participants' less than fluent levels of Spanish proficiency (LexTALE-ESP $M = 67.76\%$). If the participants had more of a range or higher proficiency scores, it is possible we would have seen Spanish proficiency have a significant effect on RT. In general, with these measures, there can be a ceiling effect. While that does not seem to be the case with the LexTALE-ESP, it could be the case with the LexTALE where the average score was quite high ($M = 89.65\%$). These issues could account for the lack of power in the models for the measures related to proficiency. Increasing the participant pool size could increase the range of proficiency scores. Additionally, having a more in-depth proficiency measure

to better tease out differences between participants in future studies seems necessary.

Another interpretation could be that having higher proficiency and/or literacy in one of the known languages when the languages are closely related (as Spanish and English are) is all that is necessary for interpreting sentences as simple as these. This view is supported by Karaca et al. (2024), who found that literacy and writing skills in *either* of their participants' languages helped with cue prediction. This could account for finding that the LexTALE, which measured the language of formal education for these participants, was a marginally significant predictor of RT rather than the LexTALE-ESP. Although the sentences were simple, the higher effect of English proficiency on the reading component of this study could be attributable to the amount of formal education in the Spanish language. Many participants did not have formal education in their heritage language and might have been forced to rely on their English proficiency or literacy. These claims, however, require further testing to be verified. Further studies could look at specifically testing participant biliteracy, including years of education in each language in the models, presenting auditory-only stimuli, and comparing unrelated languages. Transforming the study into an auditory-only study or testing literacy could control for the language-education factor. Additionally, because Spanish and English are related languages, there are a substantial number of cognates, and the grammars are similar. Thus, testing two unrelated languages could limit how much influence English proficiency has on the other language's RTs. It could also reduce the influence of shared literacy skills, especially in the case of shared orthographies, and cognates. However, the hypothesis of higher proficient participants being faster at processing Spanish OVS sentences is not supported because proficiency did not have a significant impact on RTs. Such explanations would require further study or additional analysis from a non-frequentist framework.

Finally, results from the choice RT LMEs are worth considering. In these models, similar outcomes were observed. In both languages, we found a decrease in RT in the OVS condition choices. This mirrors our previous analysis of forward transfer or interference with IA-IA sentences or good-enough processing. BLP English dominance leads to slower RT, but an increase in English proficiency leads to faster RTs in both languages. Our participants were not balanced bilinguals. In proficiency and dominance, they favored English. The participants who were more English-dominant were slower at choosing a subject as they were less able to harness the agreement cue from Spanish. However, it could be that proficiency in English was helpful in selecting the subject. LexTALE is considered an indirect measure of proficiency, under the assumption that a higher vocabulary size, what LexTALE is truly measuring, is correlated with an increase in language proficiency (Lemhöfer and Broersma, 2012; Bonvin et al., 2021). Looking past its use as a quick proficiency measure, we could interpret the findings as participants with a larger English vocabulary size are faster in an exercise with simple sentences by relying on the pure semantics of the words rather than extraneous context clues. Or the significance of the LexTALE in these models might be highlighting that bilinguals might be more apt to rely on semantic representations rather than syntactic, especially in unfamiliar or difficult syntactic constructions (Ferreira, 2003; Lim

and Christianson, 2013; Karimi and Ferreira, 2016). When we look at our other measures related to proficiency, the Flanker task was never a significant factor in our statistical models. The Flanker and LexTALE ($r = 0.22$, $p > 2.2e\text{-}16$) are poorly correlated for our participants, even more so for the LexTALE-ESP ($r = -0.063$, $p = 2.8e\text{-}06$). The absence of significant effects in the Flanker task may be attributed to the imbalance in bilingualism among our participants, who exhibited greater dominance and proficiency in English. This observation aligns with the findings of Thomas-Sunesson et al. (2018), who reported a more pronounced Flanker effect among balanced bilinguals. These results suggest the measures related to language proficiency utilized in this study may not be adequate to yield precise information about proficiency for this research context, although they are widely used in language research. To investigate this possibility further, including a separate measure of proficiency and the LexTALE to purely test vocabulary size may be appropriate. Finally, investigating heritage bilinguals from a good-enough-processing perspective more directly could reveal more clearly how cues are integrated between languages with the awareness of bilingual reliance on semantic processing.

## Conclusion

While this study was intended to further our understanding of bilingual sentence processing in the view of the UCM, due to study design and stimuli, little evidence was found corresponding to past UCM findings. According to the CM and applied studies, animacy is a weak cue in English and Spanish, in comparison to word order and agreement (Bates and MacWhinney, 1989; Pham and Ebert, 2016). However, this study suggests that the animate quality of nouns and their semantic features can prevent higher order cues from impacting how a sentence is processed. Thus, while our participants were not shown to be sensitive to the stimuli as a function of language proficiency or dominance, they do seem to be sensitive to animacy's interaction with semantics in simple sentences. Additionally, our results demonstrate that when reading Spanish sentences and choosing subjects, the non-canonical word order predicted faster RTs, due to reliance on lexico-semantics as supported by vocabulary size in English, rather than due to the syntactic frame. Overall, these findings lend support to the good-enough processing model proposed by Ferreira et al. (2002). However, it is not clear whether the shift to a good-enough processing for OVS sentences is faster for bilinguals because they do not have to search for cues supporting their syntactic representation or if it is because these bilinguals had better vocabulary than syntactic acquisition. Additional studies are necessary to tease out the relationship between inter-sentential cues and superseding factors of semantics.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by the Office of Research Integrity and Compliance, Institutional Review Board, University of California, Merced. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

RC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. ZA-M: Conceptualization, Funding acquisition, Methodology, Writing – review & editing. MS: Writing – review & editing. HB: Writing – review & editing.

## Funding

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2024.1370569/full#supplementary-material

# References

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bates, E., and MacWhinney, B. (1989). Functionalism and the competition model. *Crossling. Study Sent. Proc.* 3, 73–112.

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., and Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Devel.* 55, 341–354. doi: 10.2307/1129947

Birdsong, D., Gertken, L. M., and Amengual, M. (2012). *Bilingual language profile: An easy-to-use instrument to assess bilingualism*. COERLL, University of Texas at Austin.

Bonvin, A., Brugger, L., and Berthele, R. (2021). Lexical measures as a proxy for bilingual language dominance? *Int. Rev. Appl. Ling. Lang. Teach.* 61, 257–285. doi: 10.1515/iral-2020-0093

Brugè, L., and Brugger, G. (1996). On the Accusative 'A' in Spanish. *Probus.* 8, 1–52. doi: 10.1515/prbs.1996.8.1.1

Chrabaszcz, A., Onischik, E., and Dragoy, O. (2022). Sentence comprehension in heritage language: isomorphism, word order, and language transfer. *Second Lang. Res.* 38, 839–867. doi: 10.1177/0267658321997900

Copeland, C.-S. T. (2022). *A psycholinguistic study of individual differences in second language sentence interpretation*. The Florida State University.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cogn. Psychol.* 47, 164–203. doi: 10.1016/S0010-0285(03)00005-7

Ferreira, F., Bailey, K. G. D., and Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Curr. Direct. Psychol. Sci.* 11, 11–15. doi: 10.1111/1467-8721.00158

Frenzel, S., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2011). Conflicts in language processing: A new perspective on the N400–P600 distinction. *Neuropsychologia* 49, 574–579. doi: 10.1016/j.neuropsychologia.2010.12.003

Grundy, J. G. (2020). The effects of bilingualism on executive functions: an updated quantitative analysis. *J. Cult. Cogn. Sci.* 4, 177–199. doi: 10.1007/s41809-020-00062-5

Hernandez, A. E., Bates, E. A., and Avila, L. X. (1994). On-line sentence interpretation in spanish–english bilinguals: what does it mean to be in between? *Appl. Psychol.* 15, 417–446. doi: 10.1017/S014271640000686X

Isabelli, C. A. (2021). Contextual effects in processing OVS constructions in spanish. *Res. Second Lang. Proc. Proc. Instr.* 62:183. doi: 10.1075/sibil.62.06isa

Izura, C., Cuetos, F., and Brysbaert, M. (2014). Lextale-esp: a test to rapidly and efficiently assess the spanish vocabulary size. *Psicológica* 35, 49–66.

Karaca, F., Brouwer, S., Unsworth, S., and Huettig, F. (2021). Prediction in bilingual children. *Pred. Second Lang. Proc. Learn.* 12:115. doi: 10.1075/bpa.12.06kar

Karaca, F., Brouwer, S., Unsworth, S., and Huettig, F. (2024). Morphosyntactic predictive processing in adult heritage speakers: effects of cue availability and spoken and written language experience. *Lang. Cogn. Neurosci.* 39, 118–135. doi: 10.1080/23273798.2023.2254424

Karimi, H., and Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quart. J. Exper. Psychol.* 69, 1013–1040. doi: 10.1080/17470218.2015.1053951

Kassambara, A. (2023). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.6.0. Available online at: https://CRAN.R-project.org/package= ggpubr (accessed February 21, 2024).

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lemhöfer, K., and Broersma, M. (2012). Introducing LexTALE: a Quick and valid lexical test for advanced learners of English. *Behav. Res. Methods* 44, 325–343. doi: 10.3758/s13428-011-0146-0

Lim, J. H., and Christianson, K. (2013). Second language sentence processing in reading for comprehension and translation. *Bilingualism* 16, 518–537. doi: 10.1017/S1366728912000351

Luk, G., De Sa, E. R. I. C., and Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism* 14, 588–595. doi: 10.1017/S1366728911000010

MacWhinney, B. (2013). "The logic of the unified model," in *The Routledge Handbook of Second Language Acquisition* (London: Routledge), 229–245. doi: 10.4324/9780203808184-22

MacWhinney, B. (2022). "The competition model: past and future," in *A Life in Cognition: Studies in Cognitive Science in Honor of Csaba Pléh* 3–16. doi: 10.1007/978-3-030-66175-5_1

Mahowald, K., Diachek, E., Gibson, E., Fedorenko, E., and Futrell, R. (2023). Grammatical cues to subjecthood are redundant in a majority of simple clauses across languages. *Cognition* 241:105543. doi: 10.1016/j.cognition.2023.105543

Makowski, D., Lüdecke, D., Patil, I., Thériault, R., Ben-Shachar, M., and Wiernik, B. (2023). *Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption. CRAN*. Available online at: https://easystats. github.io/report/ (accessed February 21, 2024).

Meir, N., Parshina, O., Sekerina, I. A., and Brown, M. M. (2020). "The interaction of morphological cues in bilingual sentence processing: an eye-tracking study," in *Proceedings of the 44th Annual Boston University Conference on Language Development* (Sommerville, MA: Cascadilla Press), 376–389.

Montrul, S., and Sánchez-Walker, N. (2013). Differential object marking in child and adult Spanish heritage speakers. *Lang. Acquis.* 20, 109–132. doi: 10.1080/10489223.2013.766741

Nava, E. (2007). "Word order in bilingual spanish: Convergence and intonation strategy," in *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics* 129–139.

Özsoy, O., Çiçek, B., Özal, Z., Gagarina, N., and Sekerina, I. A. (2023). Turkish-German heritage speakers' predictive use of case: webcam-based vs. in-lab eye-tracking. *Front. Psychol.* 14:1155585. doi: 10.3389/fpsyg.2023.1155585

Pham, G., and Ebert, K. D. (2016). A longitudinal analysis of sentence interpretation in bilingual children. *Appl. Psycholing.* 37, 461–485. doi: 10.1017/S014271641500 0077

R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Available online at: https://www.R-project.org (accessed February 21, 2024).

Reyes, I., and Hernández, A. E. (2006). Sentence interpretation strategies in emergent bilingual children and adults. *Bilingualism* 9, 51–69. doi: 10.1017/S1366728905002373

Sagarra, N. (2021). "When more is better Higher L1/L2 similarity, L2 proficiency, and working memory facilitate L2 morphosyntactic processing," in *Research on Second Language Processing and Processing Instruction* (John Benjamins Publishing Company), 125–150. doi: 10.1075/sibil.62.04sag

Thomas-Sunesson, D., Hakuta, K., and Bialystok, E. (2018). Degree of bilingualism modifies executive control in Hispanic children in the USA. *Int. J. Biling. Educ. Bilingual.* 21, 197–206. doi: 10.1080/13670050.2016.1148114

Ware, A. T., Kirkovski, M., and Lum, J. A. (2020). Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Front. Psychol.* 11:1458. doi: 10.3389/fpsyg.2020.01458

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. doi: 10.1007/978-3-319-24277-4

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686