

Prediction with Code-switches: Pervasive Across Context and Experience?

1. Introduction

Code-switching, or the “alternation of two [or more] languages” within an utterance (Poplack, 1980, p. 583) is a common feature of bilingual speech and is broadly understood as encompassing and engendering multiple functions, including but not limited to identity indexation, establishment of group social norms, and the conveying of taboo or sensitive information (Appel & Muysken, 1987; Myers-Scotton, 1983). Most pertinent for this investigation is the use of code-switching (henceforth CSing) as a potential cue that upcoming input is in some way marked; that is, this proposal sketches a study of the predictive utility of CSes during online language comprehension for the anticipation of upcoming, unexpected information. In a recent paper, Tomić and Valdés Kroff (2022) argue that bilingual listeners appear to infer information about the relative frequency of an upcoming noun by using a mixed DP switch point as a signal that the following noun is unexpected (as operationalized by lexical frequency). The authors point to these results as evidence of a potential processing benefit of CSing, such that bilingual listeners may facilitate difficult-to-process input because the salience of a CS enables predictions about how (un)expected ensuing utterance content will be and allows listeners to prepare accordingly.

Critically, it is unclear from the results whether bilinguals are relying upon acquired sociopragmatic knowledge of when and where CSes are used or if they are drawing upon internal distributional statistics of CSes’ co-occurrences with low- and high-frequency words to make these predictions. One means of teasing apart these competing (though not wholly disjunctive)

hypotheses is by testing for prediction effects across *different levels of experience with CSing* and across *different kinds of speakers*, to whom listeners may ascribe different levels of sociopragmatic competence. To address these varying accounts of the driving mechanism(s) behind CS-coordinated prediction, we propose an eye-tracking study in the Visual World Paradigm (VWP) which **tests for effects of the perceived identity of a speaker and of listener experience with and exposure to CSing (as measured by the BCSP) on predictive eye-movements to low-frequency items presented in sentences containing a CS.** We begin **Section 2** by first outlining how bilingual comprehension and production of CSes have been understood in the context of predictability (2.1). We then provide a brief overview of how perceived speaker identity informs anticipatory processing and drives online comprehension along both down- and up-regulating lines, drawing primarily from research on the processing of non-native-accented speech (2.2). We round off with a discussion of recent attitudinal and quantitative work on human-AI interactions in experimental settings (2.3) before presenting our research questions and predictions in **Section 3**. **Section 4** provides a preliminary sketch of the materials and methods best suited for this investigation, while **Section 5** discusses the planned means of analysis. We conclude this proposal with a meditation in **Section 6** on the potential findings (6.1) and implications of this work (6.2).

2. Background

2.1. Code-switches as cues for upcoming unexpectedness

Research on CS processing across experimental designs (production and comprehension) and across modalities (in corpora of written and oral CSes) supports an association between CSes and informational content such that switches precede and signal heightened informational load (i.e., that input will be unexpected, with predictability inversely proportional to informational

content) (e.g., Bhattacharya & van Schijndel, 2024; Calvillo et al., 2020; Li et al., 2026; Myslín & Levy, 2015). The informational content, or surprisal, of words within utterances has been reliably shown to be directly proportional to CS likelihood across bilingual populations, with bilinguals initiating switches, for instance, when an upcoming word is less predictable given preceding content. These findings continue to be borne out, even when testing within more ecologically valid contexts and with more naturalistic data, such as spontaneous speech (e.g., Bhattacharya & van Schijndel, 2024). Broadly, however, these findings are split with respect to whether this relationship of CSing and informational content is speaker-driven, engendered by a speaker's need to more rapidly activate and access an infrequent or unpredictable word, or audience-driven and so part and parcel of a speaker's effort to facilitate her listener's comprehension. Under the former view, the co-occurrence of switches with following high-information items reflects a speaker's limited access to a low-frequency or unpredictable word in one language versus another (Calvillo et al., 2020; Sarkis & Montag, 2021), while the latter view positions the CS as a facilitatory tool directing listener's attention and cognitive resources under the impression that what will be heard or read next is likely to be costly to process (Myslín & Levy, 2015; Tomić & Valdés Kroff, 2022).

These hypotheses present potential motivations for why bilinguals engage in codeswitching: of principal interest to this study, however, is a non-orthogonal debate internal to these larger questions of the structural mechanisms promoting CSing. This debate is best articulated in Tomić and Valdés Kroff (2022), an eye-tracking study examining whether bilinguals direct more predictive fixations to an unexpected versus expected noun after hearing a switch from Spanish into English versus hearing the same sentence in Spanish only. Participants were directed to look to images using the carrier phrase “*Encuentra/Elige el ...*” (“*Choose/find*

the...) followed by either a switch into English “*drawing of a/an/Ø*” or a continuation in Spanish “*dibujo de un/una/ Ø*” and ending with either an unpredictable (low lexical frequency) or predictable (high lexical frequency) noun corresponding to one of the two images shown. Participants who heard a switch prior to an unexpected noun were more likely to fixate the target before it had been processed auditorily compared to predictive fixations in the unilingual condition. This effect was attested for both English and Spanish dominant listeners, suggesting that speakers’ knowledge of CSes’ function as a marker of the marked and unexpected is maintained irrespective of their language dominance profile. Importantly, the authors note that they are unable to disambiguate whether listeners engage in this type of prediction because comprehenders either **a).** are able to access distributional statistical information which powers inferences about what words CSes are likely to co-occur with, or **b).** are generating expectations on the basis of sociopragmatic competence developed over interaction with other CSing bilinguals which promotes an association between high surprisal words and CSes. We refer to these competing hypotheses¹ as the STATISTICAL LEARNING HYPOTHESIS and the SOCIOPRAGMATIC LEARNING HYPOTHESIS, respectively. Taking this divide as the impetus for our investigation, we ask whether it is always the case that bilingual listeners develop predictions in this precise way to a CS. By and large, this has been tacitly omitted from current research, with the exception of Tomić and Valdés Kroff (2022) and a passing consideration that listener knowledge may modulate the surprisal-CS linkage in Myslín and Levy (2015, p. 895). By testing for effects of speaker identity and listener experience, we intend to clarify this debate within CSing research. We will now briefly legitimate the Speaker Identity manipulation by drawing from research on non-native speech processing.

¹ Though the language used for these hypotheses is not explicitly used in Tomić & Valdés Kroff (2022), their underlying argumentation can be found there.

2.2. Processing in the context of speaker identity

During sentence comprehension, participants' processing of syntactic and semantic errors reflects whether the speaker is perceived as being a 'native' or 'nonnative' speaker, as indicated by ERP indices of online processing such as the N400 and P600 (e.g., Caffarra & Martin, 2019; Grey & van Hell, 2017; Hanulíková et al., 2012). Most often, perceived 'nativeness' is manipulated by presenting listeners with two distinct accents, of which one is non-normative given the participant population and context. Broadly, listeners appear to be less likely to expect L1-like production during the processing of apparently L2-accented speech and so display reduced or absent N400/P600 effects in response to grammatical anomalies, though these effects may also be due to a reduced ability to generate expectations about what they will hear next (Caffarra & Martin, 2019; Grey & van Hell, 2017; Grey et al., 2020; Hanulíková et al., 2012; but see Romero-Rivas et al., 2015, for evidence of an enhanced N400 for non-native speech). However, some studies, such as Caffarra and Martin (2019), also show that participants *do* develop *specific* expectations about the kinds of errors they are likely to encounter in non-native speech and so show selective P600 effects for uncommon errors (i.e., numbering errors) in non-native-accented Spanish versus common errors (i.e., gender-marking errors). Perceptions of speaker identity also manifestly influence lexical access, with foreign speech in priming tasks correlating with increased RTs to targets (Porretta et al., 2016). These data suggest that online language processing does not proceed uniformly given information about a speaker which informs a listener's beliefs about how ungrammatical, anomalous, or predictable input will be. More specifically, these data provide evidence that listeners may develop underspecified predictions, if they develop expectations at all, when encountering speech they believe liable to violate those expectations (Caffarra & Martin, 2019; Goslin et al., 2012; Grey & van Hell, 2017;

for adjacent work in eye-tracking finding that prediction in the L1 is suspended when errors are encountered, see Hopp, 2016). Within the context of the current investigation, we take these findings as a foundation for hypothesizing that bilinguals may adjust their expectations of how CSes are and will function when produced by a non-human agent who may be perceived as less pragmatically savvy or capable of adjusting their utterances intentionally for their interlocutor (see Section 2.3). Additionally, there is ample evidence that ERP indices absent during non-native speech processing for listeners with little to no experience with said accent are modulated by accent familiarity and a prior history of immersion in or interaction with a non-native speaker community (Caffarra & Martin, 2019; Grey & van Hell, 2017; Holt et al., 2018). In an EEG study of L1 Australian English speakers listening to Mandarin-accented English, Holt et al. (2018) found that only those participants who were familiar with the accent displayed a significant N400 to semantic violations, suggesting that experience promotes error sensitivity and poses an overall benefit for processing. Benefits of familiarity may even manifest as adaptation effects within the course of an experiment, though the polarity of effects differs from Holt et al. (2018) and others: Romero-Rivas et al. (2015) report a reduced N400 amplitude to non-native speech and interpret this component as evidence that listeners improved in their ability to process difficult, highly unexpected input. We do not comment on the diverging ERP responses in this proposal, but more generally position these studies as suggestive that similar effects may be attested in a CSing context, where beliefs about a speaker's capacity to minimize unexpectedness as part of their audience design could plausibly inform how listeners interpret strategies (such as CSing) which have been linked with signaling unexpectedness. Importantly, these beliefs may also be less specified for bilingual listeners with less overall environmental

exposure to and experience with CSing, motivating our investigation of listener CSing experience.

2.3. Human-Computer interactions as a testbed for identity effects

In order to manipulate participants' beliefs about a speaker's likelihood to signal unexpected information as part of their audience design, we intend to present listeners with CS sentences produced either by a non-human AI text-to-speech interlocutor or by a fellow human bilingual. Though some research contends that human-computer interactions proceed indiscriminately from human-human interactions (Nass & Moon, 2000; Westerman et al., 2020), especially on the basis of lexical and gestural alignment between human participants and their virtual interlocutor (Bergmann et al., 2015), comparable priming effects for passive English constructions when the prime is given by a human-like avatar versus another human (Heyselaar et al., 2017), and alignment of CSes during a Map Task between a bilingual and a chatbot (Geckt et al., 2025), there are several reasons why we should not assume human-computer interactions will reduplicate the epiphenomena of human-human interactions one-to-one. These considerations also bear on the logic of our Speaker Identity manipulation. Firstly, listeners are sensitive to whether their computer interlocutor is more or less human-like. The absence of facial features, a lack of human-like speech prosody, or an overtly 'computer-like' avatar or otherwise artificial appearance all contribute to a down-regulation of participants' likelihood to, for instance, align with the computer's lexical choices or reproduce a prime (Bergmann et al., 2015; Heyselaar et al., 2017). Additionally, participants may not ascribe intentionality to computer-produced language, such as utterances created by ChatGPT, and so tend to process more egocentrically while reporting that their computer interlocutor is less pragmatically sophisticated than a human adult (Mayn et al., 2024; Peña et al., 2023). Lastly, and specific to bilingual

speech, participants are markedly sensitive to incongruent or infelicitous mixed DP CSeS and to speech with a higher frequency of CSeS produced by a computer, with self-reported task enjoyment declining alongside performance in a Map Task in both cases (Geckl et al., 2025). We take these results as evidence that human-computer interactions may give rise to distinct processing patterns, especially for bilinguals hearing CSeS, and reason that when encountering AI-spoken CSeS, bilinguals will reflect broadly attested listener attitudes about how capable computers are of considering their interlocutor. That is, we expect that bilinguals will inhibit their expectation that CSeS are deployed strategically as a signal for upcoming unexpectedness when spoken by an AI agent versus a human bilingual.

3. Research questions and predictions

Given that (monolingual) listeners integrate information about a speaker into their online processing mechanism and so often develop less-granular expectations depending on how ‘native’ a speaker is perceived to be, this proposal extends this logic to bilingual sentence processing and asks whether CSeS maintain their predictive utility given certain information about *the CSeS*. This question partially unravels the theoretical knot presented by the dueling hypotheses about the driver(s) of CS-coordinated prediction. For instance, if listeners’ predictions of less-predictable content are influenced by how reliable they perceive a speaker to be in their ability to appropriately use a CS as an information-theoretic ‘alarm’, this would lend support to the notion that prediction using CSeS is moderated by interactional context and sociopragmatic convention. Conversely, listeners’ predictions may not be impacted by perceived speaker identity, suggesting that they may instead be relying on internal statistical and distributional knowledge that is apparently decontextualized. Thus, our first research question, **RQ1**, is as follows: to what extent are the time-course and extent of predictive eye-movements to

183 an upcoming, low-frequency target noun modulated by information about whether a speaker is
184 (un)attuned to sociopragmatic conventions of CSing?

185 We seek to answer this question by presenting bilinguals with spoken unilingual Spanish
186 and codeswitched Spanish-to-English sentences alongside naturalistic tableaux of 3 images
187 presented on a screen, of which one will be heard in the sentence (the target). Critically, one of
188 these images will represent a noun with a low lexical frequency relative to the other two images.
189 The expectation, in line with Tomić & Valdés Kroff (2022), is that listeners will direct earlier
190 and more frequent predictive visual fixations to a low-frequency noun in the Switch than in the
191 Unilingual condition. To test for whether listeners develop different processing strategies
192 depending on how knowledgeable they perceive a speaker to be of sociopragmatic and
193 information-theoretic functionalities of CSing, we will present participants with two different
194 interlocutors. All participants will hear Unilingual and CS sentences produced by either an L1
195 Spanish, L2 English bilingual or an AI text-to-speech agent. Both will be matched for a Puerto
196 Rican accent. Our core assumption is that listeners will not ascribe sociopragmatic knowledge or
197 intentionality to the speech of a non-human interlocutor and so may consequently diverge in their
198 predictive use of CSes in the AI speaker condition. That is, we anticipate listeners will not
199 perceive the AI speaker as (capable of) using CSes to mark that upcoming information is
200 unexpected and therefore predict that predictive fixations to low-frequency items in the CS
201 condition will be reduced and delayed, in line with the SOCIOPRAGMATIC LEARNING HYPOTHESIS.
202 Alternatively, listeners may not make use of speaker identity to guide their predictions at the
203 switch-site and so will continue to predict similarly across AI- and human-produced CS
204 sentences, lending support for the STATISTICAL LEARNING HYPOTHESIS.

As a complement to **RQ1**, we also aim to define the experiential factors potentially underlying CS-coordinated prediction, including but not limited to the degree of environmental exposure and personal usage of CSing a listener has. **RQ2 asks about the extent to which prediction of a low-frequency item in the Switch condition is sensitive to listener experience with and exposure to CSing.** Listeners in Tomić and Valdés Kroff (2022) reported a fairly high exposure to both oral and written CSes, though these measures were collected within the Language History Questionnaire rather than the more rigorous Bilingual Code-Switching Profile (BCSP) (Olson, 2022). Thus, the reported prediction effects may have failed to capture heterogeneous processing strategies across exposure levels which would clarify support for either working hypothesis. Specifically, if it is the case that listeners rely on statistical knowledge to guide their predictions at the switch-site (i.e., the STATISTICAL LEARNING HYPOTHESIS), the logical inference is that **listeners with more environmental exposure to CSing and higher personal usage of CSes will display more robust prediction effects relative to those with less exposure and/or experience, since the former group will nominally have more informed distributional statistics linking CSes with low-frequency words.** However, under the SOCIOPRAGMATIC LEARNING HYPOTHESIS, we might expect most listeners, regardless of their experience with CSing, to be aware of why and how CSes are used (to signal the forthcoming utterance's predictability) on the basis that even minimal experience and/or use will clarify (some of) the functions of CSes.

Favoring the SOCIOPRAGMATIC LEARNING HYPOTHESIS, we predict that all listeners, regardless of their composite score on the BCSP, will display probabilistic and temporal similarities in their predictive eye-movements to the low-frequency target in the CS condition, since we assume bilingual speakers are at least minimally aware of how CSes are used as a

salient orienting device to unexpected, high-surprisal content during production. To answer **RQ2**, we will collect objective and subjective measures of participants' relationship with CSing using the BCSP (Olson, 2022), among other questionnaires and surveys in a post-test battery. Participants' BCSP scores will then be treated as a continuous factor in subsequent analyses (see **Section 5**).

4. Methods & Materials

4.1. Participants

We plan to recruit and collect data from 120 Spanish English heritage speakers, sourced primarily from classes at a state university in the Southern United States. We will pre-screen participants along the following dimensions: must be between 18 and 35 years of age; must be a heritage speaker of Spanish; must have been raised in a predominantly English-speaking context since the age of 5; must be neurologically healthy; and must not be on brain-altering medication. Heritage speakers of Spanish were chosen due to the frequent exposure to and use of CSes within heritage speaker groups and in bilingual spaces broadly (Bullock & Toribio, 2009). This community is also heavily represented in the proposed testing locale, making data collection more efficient. The proposed number of participants was chosen given similarity to other VWP eye-tracking studies and, pending a more robust power analysis, seems generally apt for our factor design. All participants will receive compensation in the form of course credit or a cash honorarium.

4.2. Materials

4.2.1. Sentences

The proposed experiment will follow a 2 x 2 x 2 within-subject factor design, as detailed in **Table 1**. All participants will hear sentences that contain either a low-frequency or high-

frequency target noun (the Predictability condition), with frequency acting as a proxy for how unexpected a word is. Frequency data will be supplied by statistics provided by the International Picture Naming Project. Participants will also hear sentences which either contain a CS from Spanish into English (see **Table 1** for an example) or proceed unilingually in Spanish (the Language condition). Target nouns will *always* follow the switch. Lastly, all participants will hear sentences from either an AI text-to-speech agent or a human interlocutor (the Speaker Identity condition)². The human interlocutor will be a L1 Spanish, L2 English male bilingual speaking with a Puerto Rican accent. Our decision to use a Puerto Rican accent is motivated by the use of a similarly accented speaker in Tomić and Valdés Kroff (2022), though it should not be taken for granted that **perceived identity may be shaped by our choice of accent**. Still, the extent to which accent independently develops a profile of a speaker and thus informs how listeners process CSes predictively is not within the scope of this investigation. However, to mitigate potential confounding effects of speaker accent, we will match the accents of the human and AI agent interlocutors.

The AI text-to-speech agent will be sourced from Eleven Labs (2026), a subscription-based software provider that develops naturalistic AI voice packages. Currently, Eleven Labs provides 8 Puerto Rican-accented text-to-speech agents, of which 5 are male. We have opted to use the model “Berto” arbitrarily, since there are no appreciable differences across the male voices. **However, we plan to modify the text-to-speech agent by using built-in modifiers of accent similarity, stability, and exaggeration to amplify the artificial quality of the agent’s voice.** Though perhaps counterintuitive insofar as it denies a reduced contrast between the AI and

² As an important clarification in response to an anonymous reviewer, we (the experimenters) plan to create these sentences ourselves, using target nouns selected from the International Picture Naming Project. AI will not play a role in creating sentences and will only be used to auditorily present each sentence during the experiment.

human interlocutor and could potentially drive participants to ascribe different qualities to a markedly different voice (rather than a speaker whom they are simply *told* is AI), our goal with this proposal is proof-of-concept that AI-labelled speakers can spur distinct online processing behaviors, especially during prediction with CSes. Future research should aim for a tighter comparison between interlocutors. Both AI- and human-produced sentences will be normalized to 69 dB and will be controlled for length and speech rate.

		Language (within-subject)			
		CS		Unilingual Spanish	
		Speaker Identity (within-subject)			
Predictability		AI	Human	AI	Human
(within-subject)	Low-frequency	“Encuentra el drawing of a defibrillator.”	-	“Encuentra el dibujo de un desfibrilador.”	-
	High-frequency	-	“Encuentra el drawing of a pencil.”	-	“Encuentra el dibujo de un lápiz.”

Table 1. $2 \times 2 \times 2$ factor of the proposed experiment. Low-frequency target utterances for the AI condition should be assumed identical for the Human condition, and likewise for the high-frequency utterances in the Human condition; blank cells are intended to ease parsing. Target nouns bolded. Note that the target image is one of three presented. For this given example, participants would see a cat, pencil, and defibrillator.

4.2.2. Images

Images will be presented on a computer display to participants and will be arranged in triplet arrangements within a naturalistic scene. Three images, selected from open sources, will be arranged within a similarly sourced visual context, such as is shown in **Image 1**. Of the three images, one will represent a low-frequency word (e.g., *defibrillator/desfibrilador*), another a

high-frequency word (e.g., *pencil/lápiz*), and a third will act as a high-frequency distractor which participants will answer comprehension questions about (e.g., *cat*). This third image will always be an animal (e.g., cat, dog, bird). Images will also be controlled such that their Spanish translational equivalents match in gender (e.g., all Masculine or all Feminine) to prevent participants from using gender marking on the prenominal determiner for prediction (Lew-Williams & Fernald, 2007; Wicha et al., 2004). We will also norm images for visual prominence (e.g., size, color) and will present images such that a target object never appears in the same position across any two consecutive trials. Despite the complexity introduced by situating objects and corresponding interest areas into more visually complex stimuli, we emphasize the paucity of naturalistic stimuli in VWP research and argue that since prediction persists through increasing visual complexity (Heyselaar et al., 2021; Huizeling et al., 2022; but cf. Ferreira et al., 2013), VWP work should not shy away from trying to approximate naturalistic processing conditions.



Image 1. Example visual stimulus for proposed experiment. Participants would see the above while listening to the example sentence provided in **Table 1**. A potential comprehension question accompanying this image might read: “Was there a cat in the kitchen? Yes/No.” Note that images will not be AI-generated and will be extracted from open sources. Image courtesy of OpenAI (2026).

4.2.3. Stimuli Organization

We will present 132 sentences in total within one experimental block. Participants will hear 64 experimental sentences in which one of the 3 images displayed is a low-frequency noun across 8 different lists. Each experimental item will appear in one of 8 conditions across lists (Unilingual, Low-frequency Target, Human; Unilingual, Low-frequency, AI; Unilingual, High-frequency, Human; Unilingual, High-frequency, AI; CS, Low-frequency, Human; CS, Low-frequency, AI; CS, High-frequency, Human; CS, High-frequency, AI) so that participants will hear 8 sentences per experimental condition. Participants also hear 64 filler sentences which will be coupled with a visual display in which all pictured nouns are of similarly high frequencies. We plan to pseudorandomize trial order to ensure participants do not hear the same interlocutor for more than 3 consecutive trials. Prior to the experiment, participants will hear 4 unrelated practice sentences with a similar visual display to those of the main block. All sentences (filler + practice) will include comprehension questions, which will function as attention/engagement checks.

4.3. Procedure

Participants will be instructed to listen to sentences while looking at corresponding images on a screen and to answer follow-up comprehension questions. Participants will also be asked to click on the image that they hear mentioned in each sentence. Importantly, participants will also undergo a five-minute entrainment period prior to the experimental block in which they will be introduced to their interlocutors. Each interlocutor will briefly introduce themselves and read a brief (~ 2 minute) vignette in order to familiarize participants with each voice and so enable a within-subject manipulation of perceived speaker identity (for a similar design, see Trude & Brown-Schmidt, 2012; Trude et al., 2013). For the AI text-to-speech agent, the

interlocutor introduction will make explicit reference to its being an AI agent created by an unnamed software company, whereas the human interlocutor will share their name, place of birth, and occupation. Instructions for both interlocutors will contain Spanish-to-English codeswitches. Following the experimental block, participants will be asked to complete the following objective and subjective measures of bilingual contact, experience, and proficiency: English LexTALE (Lemhöfer & Broersma, 2012), LexTALE-Esp (Izura et al., 2014), adapted tests from the Diplomas of Spanish as a Foreign Language (DELE) and Michigan English Language Institute College English Test (MELICET), and the BCSP (Olson, 2022). We also ask participants to reflect on their session and comment on their perceptions of the speakers explicitly. In this debriefing questionnaire, we will ask participants about the perceived national identity of the AI speaker and how likely they would be to trust one interlocutor over the other. Likewise, participants will be asked to rate on a 5-point Likert scale how authentic and natural the CSeS produced by the AI agent were. In anticipation of participants' past exposure to AI and AI-generated speech influencing how likely those participants may be to ascribe human-like social qualities to AI agents, we will also administer a curated survey of past and current AI usage in which participants will report the extent of their personal/professional AI usage and the frequency with which they encounter AI-generated content online (e.g., on social media sites such as TikTok).

5. Analysis Plan

To answer **RQ1**, we will conduct an analysis of participants' eye-movements to target nouns using a generalized linear mixed effects model with family specified "binomial" using the lme4 package version 1.1-7 (Bates et al., 2015) in the R Environment for Statistical Computing program, version 4.5.1 (R Core Team, 2025). Looks to the target noun (the noun heard in each

experimental sentence) will be time-aligned and separated into 20 ms bins. Critically, we are interested solely in looks to the target in the window 200 ms pre- and post-target onset. Given that saccadic eye-movements take ~ 200 to 300 ms to program (Ballard et al., 1997), this window ensures we are able to capture the full breadth of predictive eye-movements. The total number of bins with looks to the target object versus the total number of bins with looks high-frequency nouns (the two competitors) will be included as the dependent variable in our GLMER. The Predictability, Language, and Speaker Identity conditions, as well as their three-way interaction, will be included as fixed effects. Participant and Item will be included as random intercepts with a maximal random effects structure as justified by the design. In the event of convergence issues, random slopes will be removed based on lowest variance explained. All factors will be deviation-coded. To answer **RQ2**, participants' BCSP score will be included as a continuous measure in the aforementioned model. We also plan to conduct a Divergence Point Analysis (Lago et al., 2023; Stone et al., 2021) to determine when predictive looks to the low-frequency target begin to diverge from those to the high-frequency target across the CS and Unilingual conditions and across the Speaker Identity conditions. This analysis will provide an onset estimate in milliseconds for when prediction of the low-frequency noun is initiated in the CS condition versus the Unilingual condition as a function of Speaker Identity.

6. Discussion & Interpretation

6.1. Predicted Findings & Alternatives

Pending concrete results, the outcomes of this study are purely conjectural. However, insofar as we can predict how bilingual listeners will behave in this experiment, we can provide a basic sketch of what certain results suggest about the use of CSes for prediction and about the dynamic mechanisms underlying bilingual speech comprehension. Firstly, if we find support for

the predictions laid out in **Section 3**, we expect to find this in the form of reduced and delayed predictive eye-movements to a low-frequency target noun in the CS condition versus the Unilingual condition when listeners hear the CS produced by an AI interlocutor rather than a human. That is, we expect to find a main effect of the Language, Speaker Identity, and Predictability, as well as an interaction of the three. Importantly, if given these effects, we would also corroborate the findings of Tomić and Valdés Kroff (2022). Likewise, we would furnish support for the Adaptive Predictability Hypothesis in demonstrating that bilinguals dynamically shift their processing strategies in response to social and experiential factors (Valdés Kroff & Dussias, 2023). That is, if bilinguals in our study begin to predict the unexpected after hearing a switch *and* if this shift in predictive behavior is modulated by perceptions of their interlocutor, then we provide convincing evidence that bilingual processing cannot be described as theoretically ‘predictable’ or deterministic.

We also expect to find that there is no effect of a participant’s experience with CSing on their predictions of an upcoming noun, such that the continuous BCSP measure does not have a main effect or interaction with the Language and Predictability conditions. However, our full effects output is unlikely to track cleanly with our predictions. For instance, it may be the case that there is an interaction of CS experience, Predictability, and Language such that more experienced CSers are less likely overall to engage in prediction of a low-frequency noun, even when presented with a CS. This would suggest that increased experience with CSing may mitigate the salience of the CS as a signal that upcoming information is unexpected or unpredictable, in turn complementing work showing that bilinguals immersed in dense CSing interactional contexts are also less likely to experience a switch-cost effect (e.g., Beatty-Martínez & Dussias, 2017; Beatty-Martínez et al., 2020), as well as work on differential processing of

CSes across levels of experience (e.g., Treadway et al., 2025; Valdés Kroff et al., 2017). That is, perhaps it is the case for these high-experience bilinguals that as CSes assume a more dominant role in everyday discourse, *both the costs and benefits* of switching decline.

Looking in the opposite direction, we may also find that it is reduced experience with CSing which corresponds with an increased likelihood and accelerated onset of predictive fixations to an unexpected object. For these bilinguals, then, the salience of the CS as an index for the unexpected may be augmented by their relative lack of experience with CSing.

Alternatively, it may be the case that the use of CSes in this fashion resembles for these less-experienced bilinguals their own tendency to use CSing as a tool for accessing low-frequency, less available items, as attested under a speaker-driven view of CSing (e.g., Calvillo et al., 2020; Montag & Sarkis, 2021). That is, less-experienced bilinguals may be more likely to engage in CS-coordinated prediction because these bilinguals themselves may CS when trying to access a low-frequency word.

We are also careful to note that we may fail to find any effect of Speaker Identity such that participants do not alter their predictive behavior when they hear an AI text-to-speech agent versus a human interlocutor. Such a finding would support burgeoning research demonstrating that language comprehension unfolds similarly across interactional contexts where the ontological status of the speaker is manipulated to be non-human (e.g., Bergmann et al., 2015; Heyselaar et al., 2017). Specifically, we point to research in predictive language processing which shows that listeners continue to anticipate objects based on verbal semantics when their interlocutor is a virtual agent in a virtual reality design (Huizeling et al., 2022), suggesting that participants may not necessarily adjust their processing strategies given this information about a speaker. Moreover, we anticipate that our planned manipulation of perceived sociopragmatic

competence via speaker identity may be confounded by prior, robust exposure to AI. Our proposed post-task survey thus aims to quantify this exposure numerically so that it may be included as a continuous factor of our models in an exploratory analysis. If participants who are more experienced with AI are more likely to engage in CS-coordinated predictions given AI-produced sentences than those with less AI experience, this would suggest that considerations of speaker identity and evaluations of sociopragmatic competence are flexible insofar as they can be qualified by experience (for similar findings, see Grey & van Hell, 2017; Holt et al., 2018). Beyond our immediate questions, this finding would also suggest that language processing research invested in questions of AI speakerhood and perceptually-conditioned processing may need to take steps to accommodate past exposure to AI into their designs and analyses.

6.2. Implications & Recommendations

Returning to this proposal's central debate, our initial predictions, absent data, broadly support the validation of the SOCIOPRAGMATIC LEARNING HYPOTHESIS. However, this should not foreclose the explanatory power of the STATISTICAL LEARNING HYPOTHESIS. For instance, if we find that speakers are more likely and faster to engage in CS-coordinated predictions under the Human speaker condition, we could argue that the difference between speaker conditions is not directly attributable to a failure to ascribe sociopragmatic knowledge to the AI agent. Rather, it may be attributable to the perceptual gulf between the two interlocutors' voices and the anomalous nature of the experimental paradigm. We might expect then that participants would show adaptation effects within the experiment across trials, such that predictive behavior converges across Speaker conditions, but only in the later stages after participants have had time to acclimate. An exploratory analysis with Time (measured in 20 ms bins) as a fixed effect in our model would be helpful in this regard. Given such behavior, it would seem then that participants'

predictive tendencies with respect to speaker identity is symptomatic of an uncertainty about whether internal statistical knowledge about CS co-occurrences is even applicable. That is, we could explain our effects within the language of the STATISTICAL LEARNING HYPOTHESIS. Similarly, for **RQ2**, an absence of an effect of CS experience could as much suggest that bilinguals need not have very developed distributional knowledge of CSes in order to predict as it suggests that all bilinguals across levels of experience have a baseline familiarity with the sociopragmatic conventions surrounding CSes.

Relatedly, it may be that the SOCIOPRAGMATIC LEARNING HYPOTHESIS is ill-formed and that bilinguals are, instead of ascribing CS-oriented sociopragmatic knowledge to a speaker, simply ascribing *rationality and the capacity for rational communicative planning* to a CSer. This is a subtle but meaningful distinction. Dating back to some of the first proto-AI technologies, such as the computer program ELIZA, there has been a well-established tendency for humans to ascribe rationality to and generally anthropomorphize their computer interlocutors (Weizenbaum, 1966; but cf. Petersen & Almor, 2025, which suggests increased AI experience moderates tendencies to anthropomorphize AI). This has also been established in the processing literature, as discussed in **Section 2.3**. Given this tendency, if bilinguals generate predictions about upcoming nouns given the presence of a CS and all that is required is that bilinguals perceive their interlocutor to be capable of rationally adjusting their speech, then it is possible we will not observe any effect of Speaker Identity. Simultaneously, if participants do not anthropomorphize the AI interlocutor but are still generating predictions under the *sole* assumption of speaker rationality, we might also expect that the predictive behavior reported in Tomić and Valdés Kroff (2022) would also hold if listeners heard CSes produced by a markedly L2 speaker of Spanish. Nominally, this speaker is unlikely to be readily ascribed the requisite

sociopragmatic ‘savviness’ to appropriately deploy CSeS, but they may be more likely to be ascribed a sense of rationality over an AI interlocutor. Thus, if listeners are sensitive to Speaker Identity but are in fact relying more on an ascription of rationality than sociopragmatic competence, the SOCIOPRAGMATIC LEARNING HYPOTHESIS as it stands misses the mark. The present proposal is not able to capture this distinction, though we recommend that future research takes up these questions.

As a final note, it should not be taken for granted that these are two *separate* hypotheses. Given that prediction writ large has been successfully argued as comprised of two stages -- one automatic and associational, the other conscious and effortful – often working in tandem with one another (see Corps et al., 2022, 2023; Nota et al., 2024; Pickering & Gambi, 2018), it seems theoretically unlikely that the two hypotheses presented herein are mutually disjunctive. Setting aside concerns over model parsimony, we may find that both hypotheses together have explanatory power. For instance, if we find that participants are more likely to engage in prediction of low-frequency nouns after hearing a CS produced by a Human than an AI *and* that more experienced CSers are more likely and faster to predict in this way (i.e., a four-way interaction of Speaker Identity, Language, Predictability, and Experience), this could be explained as experienced bilinguals drawing on both their extensive distributional knowledge and their perceptions of sociopragmatic competence to guide their predictive mechanism. Future research on bilingual adaptive predictability could begin to disentangle these accounts and clarify their relationship in a mode similar to that of research on L1 predictive processing (see Corps et al., 2022, 2023): experimental designs where distributional knowledge is taxed (our recommendation: typologically rare CSeS) or where listeners’ perceptions of speakers are stressed (our recommendation: the introduction of noise to obscure speaker identity).

491 References

- 492 Appel, R., & Muysken, P. (1987). *Language contact and bilingualism*. E. Arnold.
- 493 Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the
494 embodiment of cognition. *The Behavioral and brain sciences*, 20(4), 723–767.
495 <https://doi.org/10.1017/s0140525x97001611>
- 496 Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). “Fitting Linear Mixed-Effects Models
497 Using lme4.” *Journal of Statistical Software*, 67(1), 1–48. [doi:10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).
- 498 Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language
499 processing: Evidence from codeswitching. *Journal of Memory and Language*, 95, 173–
500 189. <https://doi.org/10.1016/j.jml.2017.04.002>
- 501 Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo,
502 R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of
503 bilingualism for language and cognition. *Journal of Experimental Psychology: Learning,*
504 *Memory, and Cognition*, 46(6), 1022–1047. <https://doi.org/10.1037/xlm0000770>
- 505 Bergmann, K., Branigan, H. P., & Kopp, S. (2015). Exploring the Alignment Space - Lexical and
506 Gestural Alignment with Real and Virtual Humans. *Frontiers in ICT*, 2.
507 <https://doi.org/10.3389/fict.2015.00007>
- 508 Bhattacharya, D., & van Schijndel, M. (2024). *Code-switching in text and speech reveals*
509 *information-theoretic audience design*. arXiv. <https://doi.org/10.48550/arXiv.2408.04596>
- 510 Bullock, B. E., & Toribio, A. J. (Eds.). (2009). *The Cambridge Handbook of Linguistic Code-*
511 *switching*. Cambridge: Cambridge University Press.

512 Caffarra, S., & Martin, C. D. (2019). Not all errors are the same: ERP sensitivity to error
513 typicality in foreign accented speech perception. *Cortex*, 116, 308–320.
514 <https://doi.org/10.1016/j.cortex.2018.03.007>

515 Calvillo, J., Fang, L., Cole, J., & Reitter, D. (2020). Surprisal Predicts Code-Switching in
516 Chinese-English Bilingual Text. *Proceedings of the 2020 Conference on Empirical*
517 *Methods in Natural Language Processing (EMNLP)*, 4029–4039.
518 <https://doi.org/10.18653/v1/2020.emnlp-main.330>

519 Corps, R. E., Brooke, C., & Pickering, M. J. (2022). Prediction involves two stages: Evidence
520 from visual-world eye-tracking. *Journal of Memory and Language*, 122, 104298.
521 <https://doi.org/10.1016/j.jml.2021.104298>

522 Corps, R. E., Liao, M., & Pickering, M. J. (2023). Evidence for two stages of prediction in non-
523 native speakers: A visual-world eye-tracking study. *Bilingualism: Language and*
524 *Cognition*, 26(1), 231–243. <https://doi.org/10.1017/S1366728922000499>

525 ElevenLabs. (2026). *ElevenLabs* [Text-to-speech software]. <https://elevenlabs.io>

526 Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world:
527 Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*,
528 69(3), 165–182. <https://doi.org/10.1016/j.jml.2013.06.001>

529 Geckt, D., Fricke, M., & Wintner, S. (2025). Strategies of code-switching in human–machine
530 dialogs. *Bilingualism: Language and Cognition*, 1–15.
531 <https://doi.org/10.1017/S1366728925100436>

532 Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent
533 processing. *Brain and Language*, 122(2), 92–102.
534 <https://doi.org/10.1016/j.bandl.2012.04.017>

535 Grey, S., Cosgrove, A. L., & van Hell, J. G. (2020). Faces with foreign accents: An event-related
 536 potential study of accented sentence comprehension. *Neuropsychologia*, 147, 107575.
 537 <https://doi.org/10.1016/j.neuropsychologia.2020.107575>

538 Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of
 539 language comprehension. *Journal of Neurolinguistics*, 42, 93–108.
 540 <https://doi.org/10.1016/j.jneuroling.2016.12.001>

541 Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When One Person's
 542 Mistake Is Another's Standard Usage: The Effect of Foreign Accent on Syntactic
 543 Processing. *Journal of Cognitive Neuroscience*, 24(4), 878–887.
 544 https://doi.org/10.1162/jocn_a_00103

545 Heyselaar, E., Hagoort, P., & Segaert, K. (2017). In dialogue with an avatar, language behavior
 546 is identical to dialogue with a human partner. *Behavior Research Methods*, 49(1), 46–60.
 547 <https://doi.org/10.3758/s13428-015-0688-7>

548 Heyselaar, E., Peeters, D., & Hagoort, P. (2021). Do we predict upcoming speech content in
 549 naturalistic environments? *Language, Cognition and Neuroscience*, 36(4), 440–461.
 550 <https://doi.org/10.1080/23273798.2020.1859568>

551 Holt, R., Kung, C., & Demuth, K. (2018). Listener characteristics modulate the semantic
 552 processing of native vs. Foreign-accented speech. *PLOS ONE*, 13(12), e0207452.
 553 <https://doi.org/10.1371/journal.pone.0207452>

554 Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language
 555 acquisition. *Second Language Research*, 32(2), 277–307.
 556 <https://doi.org/10.1177/0267658315624960>

557 Huizeling, E., Peeters, D., & Hagoort, P. (2022). Prediction of upcoming speech under fluent and
 558 disfluent conditions: Eye tracking evidence from immersive virtual reality. *Language,*
 559 *Cognition and Neuroscience*, 37(4), 481–508.
 560 <https://doi.org/10.1080/23273798.2021.1994621>

561 Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently
 562 assess the Spanish vocabulary size. *Psicológica*, 35, 49–66.

563 Lago, S., Stone, K., Oltrogge, E., & Veríssimo, J. (2023). Possessive Processing in Bilingual
 564 Comprehension. *Language Learning*, 73(3), 904–941. <https://doi.org/10.1111/lang.12556>

565 Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test
 566 for Advanced Learners of English. *Behavior Research Methods*, 44, 325–343.

567 Lew-Williams, C., & Fernald, A. (2007). Young Children Learning Spanish Make Rapid Use of
 568 Grammatical Gender in Spoken Word Recognition. *Psychological Science*, 18(3), 193–
 569 198. <https://doi.org/10.1111/j.1467-9280.2007.01871.x>

570 Li, Y., Scontras, G., & Futrell, R. (2026, March 28). Examining the effect of predictability on
 571 code-switching: A production experiment. *39th Annual Conference on Human Sentence*
 572 *Processing*. Conference on Human Sentence Processing 2026.
 573 https://hsp2026.org/abstracts/submission_46.pdf

574 Mayn, A., Loy, J. E., & Demberg, V. (n.d.). Capable but not cooperative? Perceptions of
 575 ChatGPT as a pragmatic speaker. *Proceedings of the 46th Annual Conference of the*
 576 *Cognitive Science Society*. Retrieved <https://escholarship.org/uc/item/1c381535>

577 Myers-Scotton, C. (1983). The negotiation of identities in conversation: A theory of markedness
 578 and code choice. *International Journal of the Sociology of Language*, 1983(44).
 579 <https://doi.org/10.1515/ijsl.1983.44.115>

580 Myslín, M., & Levy, R. (2015). Code-Switching and Predictability of Meaning in Discourse.
 581 *Language*, 91(4), 871–905. <https://doi.org/10.1353/lan.2015.0068>

582 Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers.
 583 *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>

584 Nota, N., Corps, R. E., Hadley, L. V., Naylor, G., & Pickering, M. J. (2024). *The effect of*
 585 *cognitive load on different stages of language prediction: A visual world eye-tracking*
 586 *study* (165,739). AsPredicted. https://aspredicted.org/3XK_YCK

587 Olson, D. J. (2023). Measuring bilingual language dominance: An examination of the reliability
 588 of the Bilingual Language Profile. *Language Testing*, 40(3), 521–547.
 589 <https://doi.org/10.1177/02655322221139162>

590 OpenAI. (2026). *Cartoon-style kitchen scene with five objects (defibrillator, pencil, mug, apple,*
 591 *and cat)* [AI-generated image]. ChatGPT Image Generator.

592 Peña, P. R., Doyle, P., Edwards, J., Garaialde, D., Rough, D., Bleakley, A., Clark, L., Henriquez,
 593 A. T., Branigan, H., Gessinger, I., & Cowan, B. R. (2023). Audience design and
 594 egocentrism in reference production during human-computer dialogue. *International*
 595 *Journal of Human-Computer Studies*, 176, 103058.
 596 <https://doi.org/10.1016/j.ijhcs.2023.103058>

597 Petersen, D., & Almor, A. (2025). Agentive linguistic framing affects responsibility assignments
 598 toward AIs and their creators. *Frontiers in Psychology*, 16, 1498958.
 599 <https://doi.org/10.3389/fpsyg.2025.1498958>

600 Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and
 601 review. *Psychological Bulletin*, 144(10), 1002–1044. <https://doi.org/10.1037/bul0000158>

602 Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL:
 603 Toward a typology of code-switching. *Linguistics*, 18(7–8).
 604 <https://doi.org/10.1515/ling.1980.18.7-8.581>

605 Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness
 606 and listener experience on word recognition. *Journal of Phonetics*, 58, 1–21.
 607 <https://doi.org/10.1016/j.wocn.2016.05.006>

608 R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation
 609 for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

610 Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to
 611 foreign-accented speech. *Frontiers in Human Neuroscience*, 9.
 612 <https://doi.org/10.3389/fnhum.2015.00167>

613 Sarkis, J. T., & Montag, J. L. (2021). The effect of lexical accessibility on Spanish-English intra-
 614 sentential codeswitching. *Memory & Cognition*, 49(1), 163–180.
 615 <https://doi.org/10.3758/s13421-020-01069-7>

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

619 Tomić, A., & Valdés Kroff, J. R. (2022). Expecting the unexpected: Code-switching as a
 620 facilitatory cue in online sentence processing. *Bilingualism: Language and Cognition*,
 621 25(1), 81–92. <https://doi.org/10.1017/S1366728921000237>

622 Treadway, H., Higdon, J., & Valdés Kroff, J. (2025). Disentangling grammar and experience: On
 623 the role of environmental exposure to Spanish-English code-switching. *Isogloss. Open*
 624 *Journal of Romance Linguistics*, 11(4), 1–27. <https://doi.org/10.5565/rev/isogloss.519>

625 Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online
 626 speech perception. *Language and Cognitive Processes*, 27(7-8), 979–
 627 1001. <https://doi.org/10.1080/01690965.2011.597153>
 628 Trude, A. M., Tremblay, A., & Brown-Schmidt, S. (2013). Limitations on adaptation to foreign
 629 accents. *Journal of Memory and Language*, 69(3), 349–367.
 630 <https://doi.org/10.1016/j.jml.2013.05.002>
 631 Valdés Kroff, J. R., & Dussias, P. E. (2023). Chapter six—Production, processing, and
 632 prediction in bilingual codeswitching. In K. D. Federmeier & J. L. Montag (Eds.),
 633 Psychology of learning and motivation (Vol. 78, pp. 195–237). Academic Press.
 634 <https://doi.org/10.1016/bs.plm.2023.02.004>
 635 Valdés Kroff, J. R., Dussias, P. E., Gerfen, C., Perrotti, L., & Bajo, M. T. (2017). Experience
 636 with code-switching modulates the use of grammatical gender during sentence
 637 processing. *Linguistic Approaches to Bilingualism*, 7(2), 163–198.
 638 <https://doi.org/10.1075/lab.15010.val>
 639 Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language
 640 communication between man and machine. *Communications of the ACM*, 9(1), 36–45.
 641 <https://doi.org/10.1145/365153.365168>
 642 Westerman, D., Edwards, A. P., Edwards, C., Luo, Z., & Spence, P. R. (2020). I-It, I-Thou, I-
 643 Robot: The Perceived Humanness of AI in Human-Machine Communication.
 644 *Communication Studies*, 71(3), 393–408.
 645 <https://doi.org/10.1080/10510974.2020.1749683>
 646 Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating Words and Their Gender: An
 647 Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and

Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*,
16(7), 1272–1288. <https://doi.org/10.1162/0898929041920487>

Response to Reviewers

The author thanks the three reviewers of this proposal for their thorough and thoughtful comments and suggestions. Responses to reviewer comments are divided into two sections below, corresponding to Minor and Major Adjustments.

Minor Adjustments

One thing that wasn't too clear to me was the sentences that are going to be read out loud the ones that will be read out loud by AI are those going to be created by the researcher or is the AI model creating the sentences in addition to speaking them

WE CLARIFIED THE SOURCE OF SENTENCES AS THE HUMAN EXPERIMENTERS (OURSELVES) RATHER THAN AI; SEE NOTE 2, P. 12.

Is there a particular reason as to why you are expecting 120 participants?

SAMPLE SIZE MOTIVATED BY SIMILAR VWP STUDIES AND PENDING POWER ANALYSIS; SEE P. 11.

p. 3, line 68: It would be helpful to provide a translation of the phrases “Encuentra / elige el...” (“Find / choose the...”).

TRANSLATION PROVIDED IN PARENTHETICAL ITALICS; SEE PP. 3-4.

p. 4, lines 81-82: Are “statistical learning hypothesis” and “sociopragmatic learning hypothesis” terms coined by the author? It may help to make this explicitly clear if this is the case.

LANGUAGE IS CLARIFIED AS OUR OWN COINAGE, THOUGH ARGUMENTATION IS ATTRIBUTED TO TOMIĆ & VALDÉS KROFF (2022); SEE NOTE 1, P. 4.

As a follow up to reading sentences out loud, your human that will be reading the sentences is a bilingual from Puerto Rico. Why from that country specifically?

OUR CHOICE OF ACCENT IS LARGELY ARBITRARY AND FOLLOWS THE EXAMPLE SET BY TOMIĆ & VALDÉS KROFF. WE DO NOT ANTICIPATE ACCENT HAVING A CONSIDERABLE INFLUENCE ON THE LIKELIHOOD OF PREDICTION, GIVEN THAT PREDICTION IS ATTESTED WHEN SPEAKERS HEAR PUERTO RICAN-ACCENTED CSES. HOWEVER, TO PARTIALLY CONTROL FOR THESE EFFECTS, WE HAVE ALSO MATCHED THIS ACCENT IN THE AI TEXT-TO-SPEECH AGENT; SEE P. 12.

it would be awesome if the researcher includes a manipulation check to confirm that participants are actually sensitive to the speaker identity condition (human vs. non-human) in the way intended.

WE HAVE INCORPORATED A POST-TASK DEBRIEFING QUESTIONNAIRE WHICH ASKS PARTICIPANTS TO RATE THE AUTHENTICITY, NATURALNESS, AND TRUSTWORTHINESS OF THE AI INTERLOCUTOR; SEE P. 16.

Major Adjustments

However, as heritage speakers navigate distinct home and societal language contexts and show great variability in dominance, there needs to be careful consideration in study design

WE INTEND TO CAPTURE SOME OF THE DYNAMICS INVOLVED IN HERITAGE SPEAKER BACKGROUNDS VIA THE INCORPORATION OF POST-TASK QUESTIONNAIRES WHICH WILL ALLOW US TO ASSESS QUESTIONS OF LANGUAGE DOMINANCE AND PROFICIENCY. HOWEVER, WE ARE HESITANT TO EMBELLISH OR EXTEND THE FOCUS ON HERITAGE SPEAKERS INsofar AS THE PRINCIPAL AIM OF THIS STUDY IS LESS AN EXPLORATION OF HERITAGE SPEAKERS' PROCESSING OF CSING AS IT IS AN EFFORT TO FIND EFFECTS OF SPEAKER IDENTITY AND CSING EXPERIENCE ON THE USE OF CSING FOR PREDICTION. FUTURE RESEARCH, SHOULD IT BE MOTIVATED, OUGHT TO EXPAND THE FOCUS TO OTHER SPEAKER COMMUNITIES (I.E., L2 SPANISH SPEAKERS OR L1 SPANISH AND L2 ENGLISH SPEAKERS).

Relatedly, participants' prior exposure to non-human speech may vary and could influence how non-human the input is perceived to be.

WE HAVE INTRODUCED A POST-TASK QUESTIONNAIRE TO QUANTIFY PARTICIPANTS' PRIOR EXPOSURE TO AI TO POWER AN EXPLORATORY ANALYSIS (P. 16).

For example, I would like to know how the researcher will operationalize lexical frequency and control across stimuli.

WE PLAN TO OPERATIONALIZE PREDICTABILITY BY MANIPULATING LEXICAL FREQUENCY OF TARGET NOUNS BASED ON AVAILABLE STATISTICS FROM THE INTERNATIONAL PICTURE NAMING PROJECT; SEE P. 12.

While the use of a non-human speaker is a fun manipulation, differences in prosody, naturalness, and processing difficulty between synthesized and human speech may unintentionally affect predictive eye-movements.

WE ACKNOWLEDGE THE POTENTIAL CONSEQUENCES OF PROSODY FOR THE UNFOLDING OF PREDICTIVE EYE-MOVEMENTS AND RECOGNIZE THAT THIS MAY CONFOUND OUR MANIPULATION OF SPEAKER IDENTITY. HOWEVER, WE AIM FOR THIS FIRST EXPERIMENT TO ACT AS A PROOF-OF-CONCEPT OF WHETHER AI VOICES LEAD TO DIFFERENT PROCESSING STRATEGIES IN THE FIRST PLACE; SHOULD DIFFERENCES FAIL TO APPEAR, FUTURE RESEARCH OUGHT TO COMPARE AN AI AND HUMAN VOICE WHICH ARE MORE SIMILAR THAN THE ONES CURRENTLY USED; SEE DISCUSSION PP. 12-3, 20-1.