Do bilinguals avoid ambiguity or minimise processing effort? Insights from an online eye-tracking study in spoken Mandarin[1]

Yajun Liu[a,1], Antonella Sorace[a,2], Kenny Smith[a,3]
[a] School of Philosophy, Psychology and Language Sciences, University of Edinburgh
[1] yajun.liu@ed.ac.uk; [2] a.sorace@ed.ac.uk; [3] kenny.smith@ed.ac.uk

Please address correspondence to:
Yajun Liu
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
Dugald Steward Building
3 Charles Street
Edinburgh EH8 9AD
yajun.liu@ed.ac.uk

[1] Acknowledgements: Following the policies of the University of Edinburgh, for the purpose of open access we have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Abstract: Previous research shows that bilinguals tend to choose more explicit referential forms (e.g., overt pronouns over null pronouns) compared to monolingual speakers, but the mechanisms driving this tendency remain debated. By conducting two experiments examining lexical ambiguity in spoken Mandarin, we tested two hypotheses:
Hypothesis 1: bilinguals would rather be redundant than ambiguous in general;
Hypothesis 2: bilinguals avoid ambiguity only when doing so helps reduce cognitive load. In Experiment 1, L1 Mandarin L2 English speakers in the UK and more-monolingual speakers in China completed a picture naming task, where we manipulated whether the context in which a picture is named makes the preferred label ambiguous (e.g., do speakers avoid saying "fen3 si1" when describing a picture of glass noodles when it appears alongside a picture of fans which shares the same label?). Experiment 2 extended this task by incorporating online eye-tracking using WebGazer. Our results showed that, contrary to Hypothesis 1, bilinguals were more likely than more-monolinguals to use ambiguous expressions. Eye-tracking analysis revealed that bilinguals tended to direct early attention toward image pairs with more accessible labels, indicating a preference for linguistic choices that are cognitively less demanding; by contrast, more-monolinguals showed proactive monitoring of ambiguity depending on their responses. These findings support Hypothesis 2 and shed light on cognitive constraints in bilingual lexical access and early signs of lexical attrition.

Keywords: bilingualism, ambiguity avoidance, processing effort, lexical ambiguity, spoken Mandarin

1 Introduction

Ambiguity occurs when a referring expression (e.g., noun or a pronoun) can refer to more than one potential referent. Controlling ambiguity so that we can still be understood by our interlocutors demands careful attention to both linguistic and contextual cues which will allow the intended referent to be recovered, or not. For speakers, the choice of a referring expression involves balancing the ease of production with the need to avoid ambiguity. For instance, reduced forms (e.g., null and overt pronouns) are generally easier to produce but can be ambiguous when the discourse context does not clearly support their interpretation, whereas more explicit forms (e.g., overt pronouns and noun phrases) are more effortful to produce but may reduce ambiguity. The Pragmatic Principles Violation Hypothesis (Lozano, 2016) suggests that bilingual speakers may prioritize clarity over economy more than their more-monolingual peers, often using more explicit forms (e.g., overt pronouns) when reduced forms (e.g., null pronouns) would be sufficient to avoid ambiguity in context. Under this hypothesis, bilinguals would be expected to rely on a more general clarity-oriented strategy, such that their preference for explicitness extends to other types of linguistic ambiguity, such as lexical ambiguity. The processing-based account (Sorace, 2019), however, suggests that bilinguals might tend to select a less cognitively demanding option (e.g., overt pronouns) when faced with processing constraints. In other words, being more explicit could either be a pragmatic failure to compute what is the most contextually appropriate referring expression, or an attempt to ease processing while making efforts to be clearer (see discussion in section 1.1). Under this account, bilinguals are less likely to avoid ambiguity when doing so is cognitively demanding, for example when it requires the speaker to monitor ambiguity closely and then select an alternative, unambiguous label.

Here we investigate these two hypotheses, exploring how L1 Mandarin L2 English (i.e. first language Mandarin, second language English) speakers in the UK, compared to their more-monolingual peers in China, handle ambiguity in spoken Mandarin, a language that is rich in lexical ambiguity at the tonal, segmental, and orthographical levels.

1.1 Bilingualism and Referential Ambiguity

The impact of bilingualism on anaphoric reference, the process of linking a referring expression (e.g., a pronoun or a noun phrase) to a previously mentioned referent, has been

extensively studied. Of particular interest are contexts where two animate referents compete for attention, as illustrated in example (1) in Mandarin (Zhang & Kwon, 2022). If a speaker here wishes to refer to Li Gang, the subject of the preceding clause, should they choose the null pronoun (given as ∅) or the overt pronoun *ta*? What if they want to refer to Wang Qiang?

1. Li Gang$_i$ gei Wang Qiang$_j$ da dianhua deshihou, ∅$_{i\,/}$ **ta**$_{i/j}$ haizai bangongshi**.**
   'When Li Gang called Wang Qiang, (he) was in the office.'

Previous research has reported a consistent tendency for bilinguals to over-accept/over-use overt pronouns in these contexts. This tendency has been demonstrated in second language learners (for Spanish L2 speakers: Margaza & Bel, 2006; Lozano & Quesada, 2023; for Italian: Belletti et al., 2007), but also child bilinguals (for Greek-English speakers in Greek: Argyri & Sorace, 2007; for Italian-English speakers in Italian: Serratrice et al., 2004; for Spanish-Italian speakers in Italian: Sorace et al., 2009), and heritage speakers (i.e, bilingual speakers who grow up exposed to a minority home language but are dominant in the majority societal language; for Mandarin: Wu, 2020; for Spanish: Montrul, 2004; for Greek: Kaltsa et al., 2015). Strikingly, it has also been demonstrated in the first language of bilinguals undergoing first language attrition: sequential bilinguals who learn their L2 after early childhood and undergo changes in their L1 due to continuous immersion in an L2 environment (for L1 Mandarin L2 English speakers: Liu, Sorace, & Smith, 2025; for L1 Italian L2 English: Tsimpli et al., 2004; for L1 Spanish L2 English: Martin-Villena, 2023). These changes in speakers undergoing attrition are particularly interesting to us, since they open a window to investigate how bilingual experience can reshape a once-stable native language, offering insights into the dynamic nature of linguistic systems and the mind's ability to adapt to evolving language experience; in the current study we focus on the native language (Mandarin Chinese) of sequential bilinguals, specifically L1 Mandarin L2 English speakers residing in the UK.

To explain the tendency towards more explicit referential forms observed in L2 learners, Lozano (2016) proposed the Pragmatic Principles Violation Hypothesis (PPVH), which attributes this tendency to a change in how bilinguals balance competing pragmatic principles. The PPVH suggests that bilinguals would rather be redundant than risk ambiguity. This hypothesis builds on the set of pragmatic principles from Grice's Cooperative Principles, particularly the Maxims of Quantity (informativeness) and Manner (clarity)

(1975). Levinson (1987a, 1987b) reformulated these maxims to account for the use of referring expressions. The Manner Principle encourages the use of a more complex or marked expression (e.g., overt pronouns or NPs) when more reduced forms (e.g., null or overt pronouns) are insufficient to resolve reference. In contrast, the Informativeness Principle encourages speakers to use minimal expressions when possible, as long as they are unambiguous. Blackwell (1998) later applied this framework to Spanish anaphora. Geluykens (2015) proposed similar ideas for English: the Clarity Principle, encouraging speakers to say as much as needed when ambiguity might otherwise arise, and the Economy Principle, which favours brevity where possible without compromising clarity. According to the PPVH, using a redundant form in topic-continuity contexts such as example 1 above (e.g., an overt pronoun instead of a null pronoun, or an NP instead of an overt pronoun) does not impair communication. This counts as a mild violation of the Informativeness or Economy Principles. However, using a more reduced form in topic-shift contexts, as when the speaker wishes to refer to Wang Qiang in Example 1 (e.g., a null pronoun instead of an overt pronoun or an NP), where ambiguity is more likely, can lead to communication breakdown, which is considered a strong violation of the Manner or Clarity Principles. While native speakers occasionally violate the Economy/Informativeness Principle, they rarely breach Clarity/Manner. Bilinguals follow a similar pattern but appear to violate the Economy/Informativeness Principle more frequently (i.e. the tendency towards overexplicitness).

Why would bilinguals exhibit different preferences when trading off clarity and economy? As the PPVH was first proposed to account for the overuse of overt pronouns and NPs in children and L2 learners, it attributes this preference to a developmental delay: the ability to avoid ambiguity emerges earlier than the ability to avoid redundancy (Shin & Smith-Cairns, 2009; Tal, Smith, Arnon, & Culbertson, 2023). However, this explanation does not readily extend to L1 attriters, who have fully acquired grammatic knowledge and pragmatic competence. Their preference for redundant forms, compared to monolinguals, is unlikely to come from incomplete acquisition. Still, the core logic of the PPVH, that bilingual speakers would rather be redundant than ambiguous, may also apply to L1 attrition. In this context, it may suggest that this motivation to avoid ambiguity may persist, not due to delayed development, but as a more general strategy shaped by bilingual experience, i.e. due to their experience of interacting with a diverse set of interlocutors in multiple languages, bilinguals

may simply be more sensitive to, or less tolerant of, potential referential ambiguity and therefore more likely to proactively resolve it.

Yet, this account faces a challenge. If overt pronouns are used to avoid ambiguity, why do bilinguals continue to use them in contexts where ambiguity remains unresolved by an overt pronoun? For example, the preference for overt over null pronouns persists in Italian when two possible referents of the same gender are present (Tsimpli et al., 2014), and also occurs in spoken Mandarin, where third-person singular pronouns "他 (he)" and "她 (she)" are phonologically identical (Liu et al., 2025). In such cases, overt pronouns do not disambiguate more effectively than null pronouns. Moreover, Liu et al. (2025) find the overuse of overt pronouns by speakers undergoing attrition even in contexts where there is only one referent and no ambiguity as to reference.

These findings seem to challenge the idea that overt pronouns are used solely to avoid ambiguity. Instead, they may suggest a more complex picture: one in which there is a distinction between what objectively disambiguates and what speakers believe disambiguates. Bilinguals may sometimes rely on forms they perceive as clearer or more informative, even when those forms do not objectively improve referential clarity. Existing literature shows that speakers do not always follow pragmatic principles in a strictly optimal way and often over-specify information. For instance, when there is only one apple in the visual context, speakers frequently describe it as “the red apple” or “the large apple”, even though the adjective is unnecessary (Engelhardt et al., 2006; Rubio-Fernandez, 2016). Similarly, in the case of pronoun use, bilingual speakers may prefer overt pronouns not because these pronouns reliably resolve ambiguity, but because they are perceived as more explicit or likely to be helpful. In this sense, bilinguals may tend to be redundant rather than risk being under-informative when faced with uncertainty. If this is the case, one would expect this tendency to generalise beyond pronoun use to other aspects of language production. This leads to our first hypothesis:

Hypothesis 1: If bilingual speakers are generally more cautious or effortful in their communication, they should be more likely than more-monolingual speakers to produce more informative utterances in other contexts involving linguistic uncertainty, such as lexical ambiguity.

We now turn to a second possible explanation that comes from the cognitive demands of pronoun usage. The overuse of overt pronouns is particularly puzzling because they do not minimize ambiguity as much as full NPs (e.g. simply repeating "Li Gang" or "Wang Qiang" in example 1) nor, surprisingly, do they merely reflect transfer from another language. Researchers have suggested that this tendency to overuse overt pronouns might be due to a general bilingual processing preference to select a cognitively less demanding option in real-time, due to fewer cognitive resources in bilinguals who need to manage more than one language in their mind (Gürel, 2019; Sorace, 2011, 2016). But why might overt pronouns be a less demanding option for bilinguals? Integrating contextual information to resolve reference, figuring out who or what the referring expression refers to and how it might be interpreted by an interlocutor, is cognitively demanding in general. Sorace (2019) argues that when processing resources are taxed, as they often are in bilinguals, speakers may default to overt pronouns, possibly because they may not be able to compute the discourse constraints that license null and overt pronouns.

The fact that speakers favour more explicit forms, even when they are not strictly necessary, may also point to more than pragmatic failure. Another possibility could be that it may reflect an effort to maintain clarity in the face of processing constraints. According to the Accessibility Theory (Ariel, 1991), referring expressions are seen as markers of referent accessibility in speakers' mental representation. For instance, reduced forms such as null pronouns are typically used for highly accessible referents, while more explicit forms signal lower accessibility. In pro-drop languages like Mandarin, subject omission is not only grammatically licensed but also a highly economical choice that minimizes production effort while assuming a shared discourse context. While dropping subjects requires minimal production effort, it imposes greater cognitive demands during planning because, for a speaker to successfully use a null pronoun, they must keep the referent highly active in their mental representation. This requires speakers to continuously track and integrate discourse information to resolve ambiguity. For bilinguals, this task may be especially challenging because managing two language systems with differing pragmatic rules for pronouns places additional strain on their cognitive resources. In contrast, overt pronouns impose slightly higher production costs, but they can serve as explicit linguistic signals that supports the continuation of a referent. As such, overt pronouns may help to reduce cognitive load while

maintaining referential clarity without resorting to a full NP, which may not be necessary if the referent is already sufficiently accessible. This leads to our second hypothesis:

Hypothesis 2: If bilinguals tend to rely on overt forms not simply to resolve ambiguity but also/instead to manage cognitive load, then they would prefer to disambiguate only when doing so is not cognitively demanding. In other words, when disambiguation requires too much effort (e.g., carefully monitoring potential ambiguity and then selecting an alternative label), they may choose not to do it.

1.2 Lexical Ambiguity

To evaluate whether the bilingual preference for explicitness reflects a broader communicative strategy of clarity over economy or whether it is conditional (i.e. modulated by cognitive load), we examine how bilinguals handle ambiguity in other linguistic domains. Like pronoun use, other referring expressions (e.g., nouns) differ in their accessibility and in-context ambiguity.

In human language, a finite set of words is used to convey an infinite range of meanings. This reuse of linguistic forms inherently introduces underspecification and ambiguity into the lexicon, as seen in phenomena such as homophones, homonyms, and heteronyms. Homophones are words that share the same pronunciation but differ in meaning (e.g., in English, *pair*, a set of two things, and *pear*, the fruit) (Trott & Bergen, 2020). A special type of homophones is homonyms, which are words with unrelated meanings which share the same pronunciation and spelling (e.g., *bat*, referring to both the animal and the baseball bat; Liang et al., 2024). Heteronyms are words that share the same spelling but differ in pronunciation and meaning (e.g., *wind*, meaning a strong breeze, or to move along a twisted path; Solomyak & Marantz, 2009).

Mandarin is rich in lexical ambiguity across three levels: tones, segments (the consonants and vowels of a syllable), and orthography (i.e., written characters). Research has extensively studied how these linguistic features contribute to lexical processing, and it has been shown that tones and segments are processed differently. For example, Sereno and Lee (2015) examined four types of prime-target phonological pairs in Mandarin in an auditory decision

task: (1) tone-and-segment overlap (ru4-ru4)[2], (2) segment-only overlap (ru3-ru4), (3) tone-only overlap (sha4-ru4), and (4) unrelated (qin1-ru4). Their results revealed the strongest priming effects occurred when both tones and segments overlapped, weaker priming when only segments overlapped, and no priming when only tones overlapped. This suggests that segments act as the primary cue for determining perceived similarity, and therefore strength of priming, while tones play a secondary role.

In Mandarin, orthography (i.e., written characters) does not directly associate with phonology (Li et al., 2022). Ambiguity stemming from orthographic features is processed differently than that from phonological cues. Qu, Li and Wei (2024) explored phonological and orthographic prediction during reading using EEG. Participants were presented with two types of sentence pairs, containing either homophonic words (as shown in example 2a & 2b) or orthographically related words (as shown in example 3a & 3b). Neural activity was measured to see whether they could predict the target word in the second sentence after reading the first one. They found that participants predicted the orthographically related target word before it appeared, whereas similarity of neural activation for homophonic words occurred only after the target word had appeared. These findings suggest that orthographic ambiguity is processed or detected more proactively than phonological ambiguity in Mandarin.

2. (a) 我们　　要　　***树木***
wo-men yao ***shu4-mu4***
'We want **trees**'
(b) 绑匪　　出价　　***数目***
bang-fei chu-jia ***shu4-mu4***
'Kidnappers demanded a ransom ***amount***.'

3. (a) 单位　　管理　　***会计***
dan-wei guan-li ***kuai4 ji4***
'Unit management ***accounting***'
(b) 上个月　　***会议***

[2] We use numbers to indicate the four tones in Mandarin.

Shang-ge-yue ***hui4 yi4***
'Last month ***meeting***'

Despite research on the comprehension of lexical ambiguity in Mandarin, relatively little is known about how native Mandarin speakers manage these ambiguities during *production*. The current study investigates lexical ambiguity in spoken Mandarin using a picture naming task adapted from Ferreira, Slevc, and Rogers (2005) and Rabagliati and Robertson (2017).

Ferreira, Slevc, and Rogers (2005) explore how native English speakers detect and resolve lexical ambiguity in referential communication. In Ferreira et al.'s experiment 1, speakers were asked to describe target objects (e.g., an animal bat) in the presence of foil objects that created *linguistic ambiguity* (e.g., a baseball *bat*). Results show that speakers did not often notice and avoid such ambiguities, often producing bare homophonic expressions for ambiguous scenes (i.e. just saying "bat"). In their experiment 2, speakers were asked to describe the target (e.g., a baseball bat) followed by a foil (e.g., an animal bat), or vice versa. In this case, speakers often produce an unambiguous expression for the second picture (e.g., saying "animal bat"). This finding suggests that while speakers may fail to proactively notice or avoid linguistic ambiguities, they are capable of detecting ambiguity in retrospect and monitoring their utterances to resolve it afterwards.

Expanding on their work, Rabagliati and Robertson (2017) examined how adults (and children) manage lexical ambiguity in referential communication, using spoken production plus eye-tracking. Their adult speakers produced more specific descriptions for linguistically ambiguous scenes than for unambiguous scenes (e.g. saying something like "animal bat" more often in a scene where a baseball bat was also present); however, this effect was much smaller than in "non-linguistic" ambiguity contexts (e.g., distinguishing between a big bat and a small bat), where visual cues like size are much more salient. Their eye-tracking data focuses on critical saccades between the target picture and the foil picture (e.g. participants' gaze shifts between the animal bat and the baseball bat) during three key phases of the task: (1) the preview phase, where participants were presented with a target picture, a foil picture and a filler picture for 4125 milliseconds; (2) the pre-naming phase, which lasted from the offset of the preview phase to the onset of participants' description of the target picture; and (3) the post-naming phase, where the three pictures were shown again for another 750 milliseconds. Their results showed that adult participants did not make more critical saccades

(e.g., between an animal bat and a baseball bat) for ambiguous scenes than for unambiguous scenes at Preview, suggesting no strong evidence for proactive monitoring of linguistic ambiguity. Although such critical saccades increased at both pre-naming and post-naming phases more for ambiguous scenes than unambiguous scenes, these effects were not statistically significant. However, their analysis of the proportions of trials that contained a critical saccade at the post-naming phase aligned with Ferreira et al (2005)'s experiment 2: adults made more saccades for ambiguous trials than for unambiguous trials, suggesting that they are likely to notice linguistic ambiguity retrospectively, after providing verbal responses. Taken together, these results suggest that linguistic ambiguity is subtle, and speakers do not often attempt to resolve it proactively.

Motivated by these findings, we conducted two experiments to address two core research questions: (1) how native Mandarin speakers manage lexical ambiguity specifically in linguistically ambiguous conditions and (2) whether L1 Mandarin L2 English bilingual speakers avoid ambiguity to a greater extent than their more-monolingual counterparts, potentially reflecting a broader strategy of ambiguity avoidance in bilinguals. In both experiments, we recruited two groups of participants: more-monolingual Mandarin speakers in China and L1 Mandarin L2 English bilingual speakers in the UK. Participants completed a picture naming task for both experiments; Experiment 2 builds on Experiment 1 by incorporating online eye-tracking to capture participants' gaze behaviours during the naming process.

2 Experiment 1

2.1 Data Availability

The data (including participants' responses, experimental stimuli and full analysis) that support the findings of this study are openly available on the Open Science Framework at https://osf.io/kc3tr.

2.2 Method

Participants completed a picture naming task in their L2, spoken Mandarin, adapted from Rabagliati & Robertson (2017), then completed a questionnaire assessing their use of and exposure to Mandarin and English.

Participants

Twenty-four Mandarin-English bilinguals in the UK and 23 more-monolingual speakers in China took part in the experiment. All participants were university students. The bilingual participants had stayed in the UK for 12-84 months (Mean: 35.54 months, SD: 18.27) at the time of participation and were aged 19-29 years old (Mean: 24.21 years, SD: 3.11). They began learning English between the ages of 4 and 12 years old (Mean: 6.29, SD: 1.83). The more-monolingual participants, who had never been abroad, were aged 23-33 years old (Mean: 26.57 years, SD: 2.71). Teaching of English is standard in the education system in China (as is, for example, some teaching of French, German or Spanish for all children in the UK) and so our more-monolingual participants also had some experience with English; they began learning English between the ages of 3 and 12 years old (Mean: 7.91 years, SD: 2.09).

In their questionnaire responses, bilingual participants reported higher English proficiency than the more-monolingual participants across four language skills (speaking, listening, reading, and writing), as well as greater English use across the same four skills and 12 daily contexts (e.g., at school, with roommates or neighbours, with friends, during social events or activities, while shopping, reading, emailing, texting, using social media, or watching shows). Figures 1-3 in Appendix A provide descriptive statistics on these questionnaire responses.

Stimuli

We selected pictures whose descriptions in spoken Mandarin will result in potential lexical ambiguity at the tonal, segmental, and orthographical levels. We tested four categories of lexical ambiguity: homonymy (complete overlap of segments and orthography between members of the pair), tone-and-segment overlap (complete overlap of spoken form but different orthography), segment-only overlap (spoken form overlaps in segments but not tones, different orthography), and first-character-only overlap (no overlap in spoken form, first written characters share the same form). Table 1 presents examples of word pairs in these categories. Each word pair is illustrated with two distinct pictures corresponding to the two distinct meanings.

Table 1: *Examples of word pairs in four categories.*

Category	Spoken Form	Orthographic Form	Meaning

Homonymy	*fen3 si1*	粉丝	fans
	fen3 si1	粉丝	glass noodles
Tone-and-Segment Overlap	*shou3 shi4*	首饰	jewelry
	shou3 shi4	手势	hand gestures
Segment-Only Overlap	*hua1 ban4*	花瓣	petals
	hua2 ban3	滑板	skateboard
First-Character-Only Overlap	*bo4 he2*	薄荷	mint
	bao2 bing3	薄饼	thin wrap

On all trials participants are shown 3 pictures and asked to name one of them, the target. On *ambiguous* trials, the picture array consists of the target, a competitor, and a filler (see Figure 1), where two members of a word pair are used as the target and competitor, creating potential referential ambiguity (e.g. in a homonymy pair, the target image might be of fans, the competitor image of glass noodles, and the filler an unrelated image, e.g. glasses). In *unambiguous* trials, one picture from a word pair is selected as the target picture, shown alongside a new filler image (filler 1) and the same filler image that appeared in the corresponding ambiguous trial (filler 2). This design is consistent with the method used in Rabagliati and Robertson (2017). For each target picture, the *target label* (a Mandarin expression as seen in Table 5.1) corresponds to the lexical item in the word pair that the picture is designed to elicit. In unambiguous trials, the target label clearly refers to a single image. In ambiguous trials, the same label is either identical to, or similar to, the label for another picture in the array, creating potential ambiguity through homonymy, phonological overlap, or orthographic similarity.

Each participant is tested on 32 sets of pictures, randomly selected from an inventory of 160 sets of pictures (8 sets for each of the four ambiguity types in total, 4 in ambiguous trials, 4 in unambiguous trials). If a target picture is selected for an ambiguous trial for a given participant, it will not be shown in an unambiguous trial for that participant. This ensures that participants are less likely to detect the underlying design of the experiment or be primed on the potential ambiguity of target images that recur across trials.

Procedure

The experiment was built with jsPsych (De Leeuw et al., 2023) and conducted online with an accompanying Teams or Tencent video call (Tencent is a video calling platform similar to Teams widely used in China) . Participants completed the experiment in their web browser while sharing their screen with the researcher, to ensure that they understood the instructions

and performed the task correctly. After completing the experimental task, they were asked to complete the questionnaire.

The procedure of the picture naming task followed the production experiment in Rabagliati and Robertson (2017) (see Figure 1). Participants first saw three pictures for 4125 milliseconds (Preview Phase). The positions of pictures (target, competitor and filler on ambiguous trials; target and 2 fillers on unambiguous trials) were randomized. This was followed by the Pre-Naming Phase where a Mickey Mouse icon appeared next to the target picture. Participants clicked a microphone button located below the pictures to start and stop recording their description for Mickey Mouse using one Mandarin expression (Naming Phase). Then, they proceeded to the Post-Naming Phase, during which the three pictures were displayed for a further 4125 milliseconds.

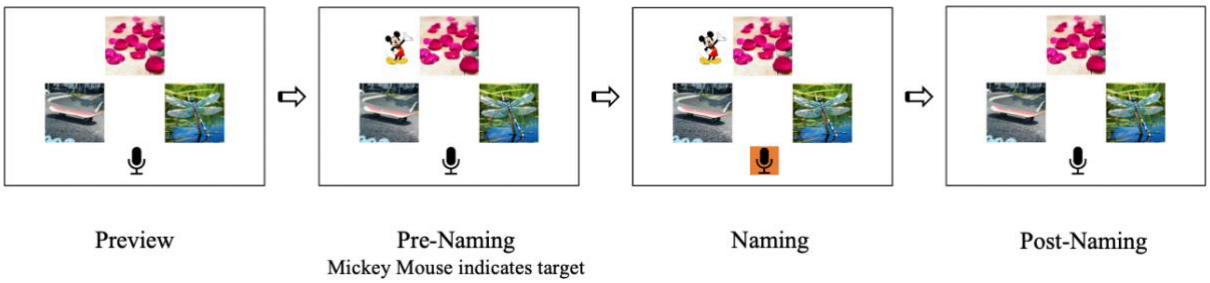

Figure 5.1: An example of the Segment-Only Overlap category in Experiment 1: target is *hua1 ban4* (*花瓣*), petals, competitor is *hua2 ban3 (滑板),* skateboard, filler is *qing1 ting2 (蜻蜓),* dragonfly.

2.3 Predictions

Our experimental design allows us to examine whether speakers tend to avoid ambiguity in their verbal responses. We predict that if speakers generally tend to avoid ambiguity, they will have more target-label responses in the unambiguous condition than in the ambiguous condition (e.g., they will say "fen3 si1" to describe the picture of fans in the unambiguous condition; in the ambiguous condition, they may produce more elaborated or self-corrected expressions, i.e. "ming2 xing1 fen3 si1" [fans of a celebrity] instead). We expect fine-grained differences among the four categories of lexical ambiguity. Homonymy pairs represent the most ambiguous scenes as they overlap in all three levels, whereas the first-character-only overlap pairs might be the least ambiguous as the two labels partially share orthographic form but are pronounced completely differently, making the "ambiguity" even more subtle in spoken language (but recall that Qu et al. 2024 did find effects of this kind of ambiguity). We

expect speakers to avoid ambiguity more for homonymous pairs. If bilingual speakers tend to avoid ambiguity to a greater degree than their more-monolingual peers, we expect them to have a stronger tendency towards avoiding ambiguous expressions in the ambiguous condition.

2.4 Data Analysis

We focused on participants' verbal responses and used a different coding scheme from Rabagliati and Robertson (2017). In their study, specific descriptions that can only be applied to the target picture (e.g., small dog, dog on the left) were coded as target descriptions and all the other descriptions as non-target. In our case, such elaborated expressions containing the target label (e.g., "ming2 xing1 fen3 si1" [fans of a celebrity] for the image of fans) accounted for less than 10% of the trials. This difference may come from the design of our task. We examined four types of lexical ambiguity, and many of our images were visually complex or open to multiple interpretations. For example, when describing the image of fans, some participants used phrases like "ying4 yuan2 tuan2" (meaning "supporters' group"), which are contextually appropriate but not applicable to the competitor image (i.e. glass noodles, also fen3 si1). However, such specific descriptions did not clearly indicate whether speakers were deliberately avoiding the ambiguous labels or simply unaware of the potential ambiguity.

Therefore, in our analysis, if the produced expression exactly matched the target label, it was coded as a target description; otherwise, it was coded as a non-target description. This approach allowed us to test our hypothesis in a more straightforward way: If participants produced fewer target labels in the ambiguous condition, it would likely indicate they were avoiding ambiguity.

Six trials were excluded from data analysis, including three empty responses and three responses that were either unintelligible or unrelated to the picture being described. As a result, 748 trials in the ambiguous condition and 750 trials in the unambiguous condition were analysed. Bayesian mixed-effects logistic regression was used to examine the log-odds of producing targe descriptions. We also examined speakers' word responses across the four ambiguity types by fitting a separate logistic regression model.

Both analyses were conducted using the brms package (Bürkner, 2017) in R (R Core Team, 2024). The probability of direction (pd) was computed using the pd function from the bayestestR package (Makowski et al., 2019). For each model, we used very weakly informative priors with mean 0 and standard deviation 1.5 (log-odds) for both the intercept and the other effects, corresponding to a 95% Credible Intervals between -3 and +3 log-odds, equal to almost 0 to 100% probability). Four MCMC chains of 4000 iterations each were executed and the first 1000 iterations were warmup.

2.4 Results

Verbal responses

Figure 2 illustrates the use of participants' target-label responses across conditions. Participants produced the target labels our image stimuli were designed to elicit around half the time, with the remaining responses being other descriptions which were compatible with the target image but revealing a different conceptualization of the object (e.g., saying "xing1 kong1" (star sky) for the image of stars). Our first analysis of verbal responses predicts target-label responses (coded as "1") versus non-target responses (i.e. alternative, unambiguous expressions, coded as "0"). The model included fixed effects of Condition (unambiguous or ambiguous trial), Group (more-monolingual or bilingual), and their interaction. We used the default treatment coding for fixed effects of Condition and Group, with the *ambiguous* condition and *more-monolingual* group as reference levels, respectively. Random intercepts for participants and items (each 3-image array treated as a separate item), as well as by-participant random slopes for Condition and by-item random slopes for Group, were included.

The analysis shows that monolinguals did not differ from bilinguals in the ambiguous condition (b = 0.25, CrI = [-0.28, 0.79], pd = 82%), although bilinguals numerically tended to use more target labels for ambiguous trials. More-monolinguals did not reliably use more target labels for unambiguous trials, compared to ambiguous trials, as indicated by the wide credible intervals for the effect of Condition indicate (encompassing 0: b = 0.16, CrI = [-0.56, 0.88], pd = 67%). There was no reliable interaction between group and condition (b = -0.18, CrI = [-0.75, 0.39], pd = 74%), suggesting that the effect of ambiguity did not statistically differ between groups. Full results are presented in Table 1 in Appendix B.

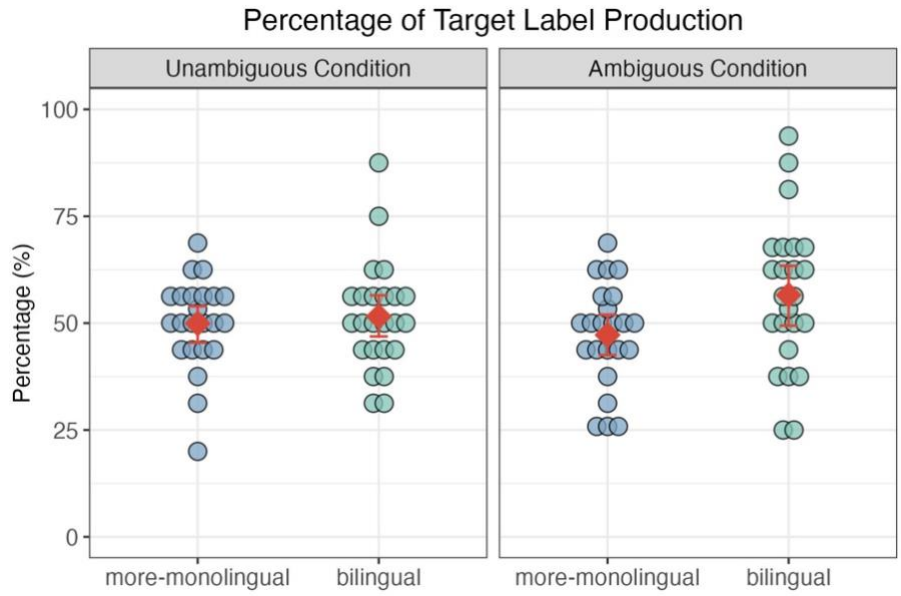

Figure 2: The use of target labels in the unambiguous and ambiguous conditions in Experiment 1. Each dot represents the % of target labels produced by one speaker; the red diamond indicates the mean of by-participant percentage, with error bars showing bootstrapped 95% confidence intervals of the mean.

Figure 3 shows the use of target labels across four ambiguity types. Our second model included an additional fixed effect of Ambiguity Category, along with the three-way interaction of Condition, Group, and Ambiguity Category. As in the first model, Condition and Group were treatment-coded with the same reference levels. Successive difference contrasts were applied to Ambiguity Category since it had an ordinal structure from most to least ambiguous type. This contrast coding enables comparisons between adjacent levels of ambiguity: tone-and-segment overlap vs homonymy, segment-only overlap vs tone-and-segment overlap, and first-character-only overlap vs segment-only overlap. The random-effects structure included random intercepts for participants and items, by-participant random slopes for Condition, Ambiguity Category, and their interactions, as well as by-item random slopes for Group.

The results show that, in the Ambiguous Condition, more-monolinguals reliably used more target labels for tone-and-segment overlap pairs than for homonymy pairs (b = 1.52, CrI = [0.33, 2.76], pd = 99%). The differences across the other two contrasts were not reliable as the credible intervals included 0: segment-only overlap vs tone-and-segment overlap (b = 0.92, CrI = [-0.13, 2.76], pd = 96%); first-character-only overlap vs segment-only overlap (b = -0.97, [-2.06, 0.12], pd = 96%). We expected the largest difference between conditions and/or groups to emerge for homonymy pairs, given their complete overlap at all three levels.

However, no reliable interaction effects were observed, suggesting that these patterns broadly held for unambiguous trials and for bilingual speakers. The largest numerical difference between groups appeared in the first-character-only overlap category in ambiguous condition, as seen in Figure 3. Table 2 in Appendix B presents the full results.

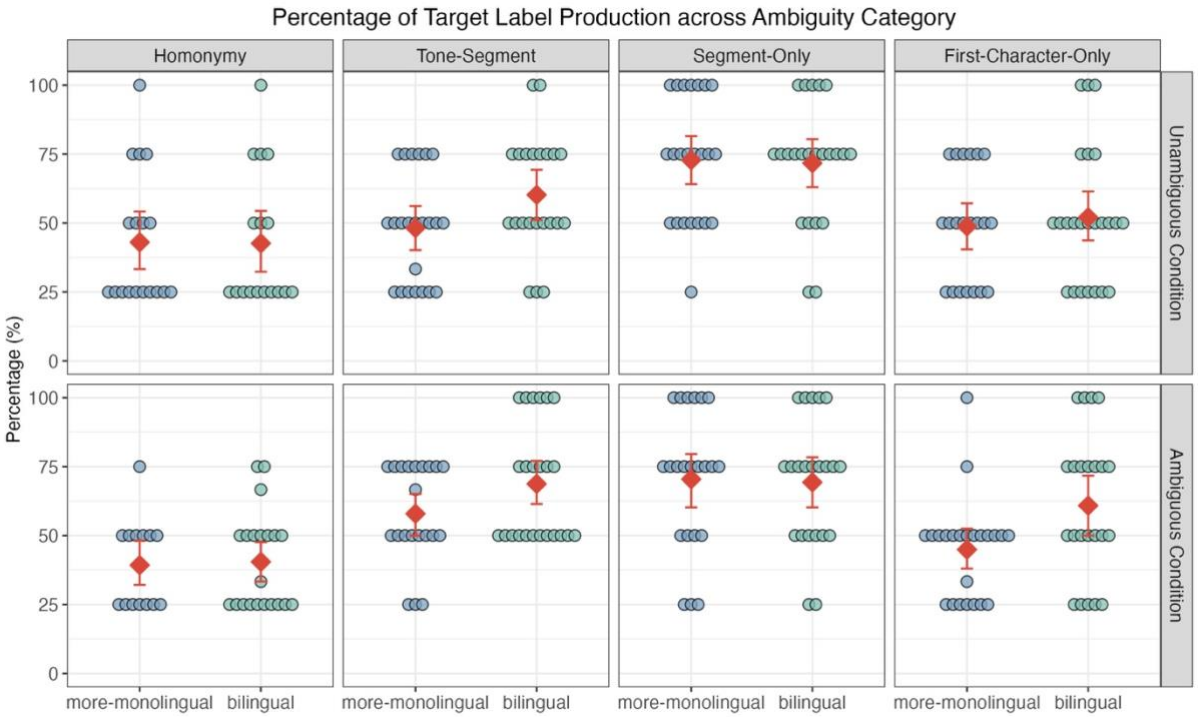

Figure 3: The use of target labels across conditions and ambiguity types in Experiment 1. Plotting conventions as in Figure 2.

2.5 Discussion

Neither group consistently avoided ambiguity by producing more unambiguous expressions in the ambiguous condition (overall or across ambiguity types) and no reliable group differences were observed. Homonymy pairs were expected to elicit the strongest ambiguity due to full overlap at all three linguistic levels; they showed the lowest rate of target description in both the ambiguous and unambiguous conditions, suggesting that the low target production for this category may be driven by specific properties of those items themselves, rather than the ambiguity. For example, in the homonymy pair *du4 juan1*, 杜鹃, which can refer to either a type of bird (cuckoo) or flower (azalea or rhododendron), the target referents do not have highly distinctive or widely recognizable visual features. As such, the actual images used may not have clearly evoked these specific meanings for participants, leading to relatively low overall rates of target phrase production. However, the fact that there are no reliable differences

between ambiguous and unambiguous trials indicate that our speakers are quite insensitive to these types of ambiguity, at least in their verbal productions, consistent with prior findings that linguistic ambiguity may be too subtle to trigger proactive disambiguation.

The lack of group effects in our data (i.e. no clear difference between more-monolingual and bilingual participants) do not support the hypothesis that bilinguals prioritize clarity over economy in general. If anything, bilinguals numerically used a higher percentage of targe-label responses in the ambiguous condition, compared to the unambiguous condition and to monolinguals. It may be that bilinguals are more sensitive to the ambiguity associated with pronouns, but this sensitivity does not extend to other types of referring expressions; alternatively, bilinguals' overexplicitness in pronominal reference might be driven by other factors. At the same time, more-monolinguals themselves did not significantly disambiguate in the condition where potential lexical ambiguity existed; therefore, we cannot attribute the absence of group difference to a cognitive burden on bilinguals given current findings.

Despite the null effects, the results still reveal patterns worth further exploration. If the tendency for bilinguals to produce more target labels in the ambiguous condition were to prove robust, our Hypothesis 2 could be considered as a possible explanation: the competitor's overlapping features (i.e. tone, segment, or orthography) co-activates the same lexical entry as the target, making the shared label more readily accessible; bilinguals, who already face higher processing demands due to managing two language systems, may be more likely to retrieve this highly activated (ambiguous) label, reducing the cognitive load to search for an alternative description.

In Experiment 2, we replicate Experiment 1 but incorporate online eye-tracking to investigate speakers' gaze shifts between the target and competitor images, following Rabagliati and Robertson (2017). For instance, in cases where participants avoid producing target labels, it remains unclear whether speakers deliberately avoided ambiguity or simply failed to notice the potential ambiguity.

3 Experiment 2

In this experiment, participants completed the same web-based picture naming task in spoken Mandarin (with modifications in visual layout), with their gaze data collected via their

webcam using the jsPsych WebGazer plugin (Papoutsaki et al., 2016). Prior to the main experiment, all participants completed two eye-tracking calibration tasks to ensure a reasonable accuracy of gaze tracking. Only those who passed both tasks were invited to continue. In the main session, they repeated both calibration tasks and then completed the picture naming task, before finally filling out the same language-use questionnaire used in Experiment 1.

3.1 Methods

Participants

Fifty-eight L1 Mandarin L2 English bilingual speakers residing in the UK and 67 more-monolingual Mandarin speakers in China initially completed two calibration tasks. All of them were university students. Of these, 45 bilinguals and 49 more-monolinguals passed both eye-tracking calibration tasks and proceeded to complete the main session.

Among the 45 bilingual participant, one was excluded from all analyses due to too many missing audio recordings from the picture naming task. Ten further participants failed one or both calibration tasks during the second calibration session (immediately preceding the picture naming task), and two passed but had too many invalid fixations (e.g., more than 30% fixations excluded for being off-screen) during the picture naming task. A further three participants lacked on-screen image-order information from the picture naming task due to a coding error in the construction of the experiment. Without this information, we could not map their fixations to the corresponding regions of interest (ROIs). As a result, these 16 participants were excluded from the eye-tracking data analysis (leaving us with a sample of 29 participants for the eye-tracking analysis), but their data were included in the analysis of verbal responses, giving us 44 bilingual participants in the verbal response analysis. These 44 participants' ages ranged from 22 to 35 years (Mean: 25.93 years; SD: 3.38). They began learning English between the ages of 3 and 13 (Mean: 7.36 years; SD: 2.81) and had been living in the UK for six to 69 months (Mean: 19.66 months; SD: 17.97).

Among the 49 more-monolingual speakers, eight were excluded from all analyses: three did not complete the main experiment due to technical problems (e.g., non-functional built-in microphones), one had too many implausible verbal responses, and four reported too high English use in their daily life, making them unsuitable for inclusion in the more-monolingual

group. Another five participants failed one or both calibration tasks during the second calibration session. Consequently, 41 more-monolingual speakers were included in the verbal response analysis. Their age ranged from 21 to 33 years old (Mean: 24.22 years; SD: 2.94). They began learning English between the ages of 3 and 13 years old (Mean: 8.24 years; SD: 2.27) and had never been abroad at the time of participation. Thirty-six of these participants were included in the eye-tracking analysis.

As in Experiment 1, bilingual participants in this experiment reported higher English proficiency across the four language skills and greater English use across a range of daily contexts compared to the more-monolingual group (see figures 4-6 in Appendix A).

Stimuli and procedure

The same picture stimuli from Experiment 1 were used in the picture naming task, but the visual layout was adjusted to enhance gaze tracking accuracy. Specifically, a quadrant-based design was implemented (see Figure 4), as previous research has shown that WebGazer performs more reliably in such a layout (Slim & Hartsuiker, 2023). The screen was divided into four quadrants. On each trial, three images (e.g., target, competitor, and filler) were displayed in three of the four quadrants randomly, with one quadrant remaining blank. Each image was centred within its respective quadrant.

As in Experiment 1, this task had four phases (Preview, Pre-Naming, Naming, Post-Naming). One key modification in the current experiment was the recording procedure: instead of clicking a microphone button on the screen, participants were instructed to press the space bar to start and stop recording. After pressing the space bar, the background of the Mickey Mouse turned orange, indicating that recording had begun and that participants could begin speaking. This modification was intended to reduce attentional shifts away from the images, as participants no longer needed to visually locate and click a button on the screen to initiate recording.

Participants were also asked to complete two calibration tasks prior to the main task: (1) a dot-based task and (2) a quadrant-based task. The dot-based task consisted of two steps, which together formed one calibration attempt. In Step 1, a black dot (12 pixels in diameter) randomly appeared at one of the five locations on the screen (centre and four corners), twice at each location, 10 trials in total. Participants were instructed to click the dot each time it

appeared. At Step 2, the dot again appeared randomly at the same five locations, twice at each location, but this time participants were instructed to look at the dot and follow it with their eyes as it moved. Calibration accuracy of this task was assessed by comparing gaze coordinates to the dot locations in Step 2. Participants passed if at least 50% of gaze samples landed within the 250-pixel ROI at each dot location. Each participant completed a minimum of two calibration attempts. If they met the accuracy criterion within those first two attempts, they moved on to the next calibration task; otherwise, they continued until they passed, with a maximum of five attempts in total.

The quadrant-based task used the same four-quadrant layout as the picture naming task. On each trial, the Chinese phrase "看这里", meaning "Look here", appeared at one of the four quadrants on the screen each time. Each quadrant displayed the phrase three times in a random order (12 trials in total). Participants were instructed to look at the phase and follow it with their eyes when it moved. They passed if, on average, at least 75% of fixation samples fell within target quadrants.

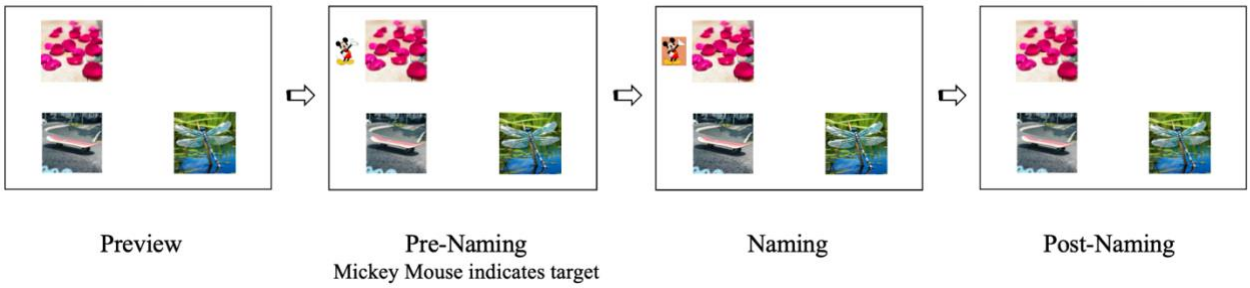

Figure 4: An example of the four-quadrant visual display in Experiment 2.

3.2 Predictions

Findings from Experiment 1 suggested that bilinguals did not avoid lexical ambiguity to a greater extent than more-monolinguals. Albeit statistically not robust, bilingual participants numerically used more ambiguous expressions in the Ambiguous Condition. This trend motivates a shift in focus on Hypothesis 2: bilinguals may not avoid ambiguity when doing so requires more cognitive effort.

In terms of eye movement, we predict that if participants detect the potential lexical ambiguity in the Ambiguous Condition, they may make more saccades between the target

and competitor images. Prior findings (from Ferreira et al., 2005, and Rabagliati & Robertson, 2017) have shown that speakers often monitor their utterances retrospectively. Therefore, we expect more saccades particularly at the Post-Naming Phase. For instance, if bilinguals do not resolve ambiguity in speech (i.e. they choose the default, ambiguous label more than more-monolingual speakers), we expect them to show even more saccades at this phase, reflecting delayed awareness or retrospective monitoring of the ambiguity after speaking.

3.3 Data Analysis

Verbal response

Verbal responses of 44 bilingual and 41 more-monolingual participants in Experiment 2 were analysed using the same method as in Experiment 1. Eleven trials were excluded from analysis, including nine empty responses and two unintelligible or unrelated responses. One participant was missing 4 trials. Consequently, a total of 1351 ambiguous trials and 1354 unambiguous trials were analysed. As in Experiment 1, we fitted two Bayesian mixed-effects logistic regression models with the same by-participant and by-item random intercepts and slopes to examine expression production.

Eye-tracking calibration

For eye-tracking data from the 29 bilingual and 36 more-monolingual participants who passed both calibration sessions, we first examined their calibration accuracy in the dot-based and quadrant-based tasks. Separate analyses were conducted for the two calibration tasks to assess whether baseline eye-tracking performance differed between groups. This step was intended to ensure that any group-level differences observed in the main picture naming task could not be attributed to discrepancies in calibration quality.

For the dot-based task, we used the default implementation of the WebGazer plugin in jsPsych, which recorded accuracy percentages (%) for each dot position rather than raw fixation coordinates. Accordingly, Welch's t-tests were conducted on the accuracy data using the t.test function in R between groups in each calibration session.

For the quadrant-based calibration task, we saved individual fixation coordinates. As such, a Bayesian logistic mix-effects regression model was fitted to examine group differences

across two calibration sessions. In this task, the entire quadrant in which each image appeared served as its ROI, following conventions in visual world paradigm studies (e.g., Dijkgraaf et al., 2017; Slim & Hartsuiker, 2022). The broader ROI was intended to account for fixations that may not land exactly on the image but still show attention to it. Therefore, fixations within the target quadrant were coded as 1, while those that fell outside the display area or exactly on any boundary lines (the interval lines dividing the screen into four quadrants and the edges of the screen) were treated as invalid fixations and coded as 0. Due to the large number of observations (192,321 individual fixation points), this model was run with four MCMC chains of 2000 iterations each (default settings), rather than 4000 as in our other models, with the first 1000 iterations as warmup.

Saccades in picture naming

In the picture naming task, we used the same ROI boundaries and fixation criteria as in the quadrant-based calibration task. We excluded invalid fixation samples and single sampled fixations that were not in a sequence of at least two fixations to the same quadrant. The exclusion rate is 6% for the bilingual group and 3.4% for the more-monolingual group.

Our eye-tracking data analysis followed the approach in Rabagliati and Robertson (2017). We first assigned fixations to the Pre-Naming or Post-Naming phase based on when participants started speaking, rather than when they pressed the space bar to start/stop recording; because our picture naming task was self-paced, participants often delayed speaking after pressing the space bar to start recording or paused before pressing it again to stop. To accurately define the pre-naming and post-naming phases, we manually annotated speech onset and offset using Praat version 6.4.23 (Boersma & Weenink, 2024), based on acoustic features such as amplitude and pitch. From these annotations, we extracted the before-speech delay (the duration between the start of recording and speech onset), the after-speech delay (the duration between speech offset and the end of the recording), and the total duration of each recording. We then used these delay durations to reassign fixations that were automatically labelled as part of the Naming Phase. Specifically, fixations with timestamps earlier than the before-speech delay were reassigned to the Pre-Naming Phase, and those with timestamps later than the point marking the start of the after-speech delay (i.e. total duration minus after-speech delay) were reassigned to the Post-Naming Phase. Fixations that did not fall within either delay window remained in the Naming Phase.

Next, we identified saccades as gaze shifts between image quadrants and excluded any saccades involving the blank quadrant. In the Ambiguous Condition, saccades were classified as *target-competitor*, *target-filler*, and *competitor-filler*. In the Unambiguous Condition, saccade types were *target-filler1*, *target-filler2*, *filler1-filler2*. Saccades during the Naming Phase were excluded from analysis because the time window was relatively brief, and many participants did not make any gaze shifts during this period. This left 19,946 saccades for analysis. Following the analysis in Rabagliati and Robertson (2017), *target-competitor* saccades were defined as the critical saccade type in the Ambiguous Condition, while *target-filler1* saccades were used as the critical type in the Unambiguous Condition for comparison. Those critical saccades were coded as 1, and all the other saccade types were coded as 0. Unlike Rabagliati and Robertson (2017), who fitted separate models for each phase (preview, pre-naming, post-naming), we fitted a single Bayesian mixed-effects logistic regression model that included Phase, Condition, Group, Response Type (target or non-target label produced), as well as their interactions.

All models used the same weakly informative priors and four MCMC chains as in Experiment 1 (Mean = 0, SD = 1.5 log-odds, 4000 iterations per chain), unless otherwise specified.

3.4 Results

We begin this section by presenting the statistical results on participants' verbal responses in the picture naming task. This is followed by statistical analysis of online eye-tracking calibration accuracy for participants who passed the screening criteria during both calibration sessions. Lastly, we analyse their looking behaviours (i.e. saccades) across different experimental phases and conditions in the picture naming task.

Verbal responses

Figure 5 illustrates the use of participants' target-label responses across conditions. Participants in the current experiment also used the targets labels about half the time. Our first analysis shows that more-monolingual speakers tended to use more target labels in the unambiguous condition than in the ambiguous condition, but this tendency was not reliable as the wide credible intervals include 0 (b = 0.23, CrI = [-0.47, 0.97], pd = 73%). However, bilingual used more target labels than more-monolinguals in the ambiguous condition (b =

0.44, CrI = [0.06, 0.82], pd = 99%), consistent with Experiment 1. Additionally, there was a reliable interaction effect between group and condition, indicating the different effects of trial on the two groups: while the more monolingual group uses numerically more target descriptions in the unambiguous condition relative to the ambiguous condition, bilingual speakers used *fewer* target labels in the unambiguous condition (b = -0.47, CrI = [-0.88, -0.05], pd = 99%). Full results are reported in Table 3 in Appendix B.

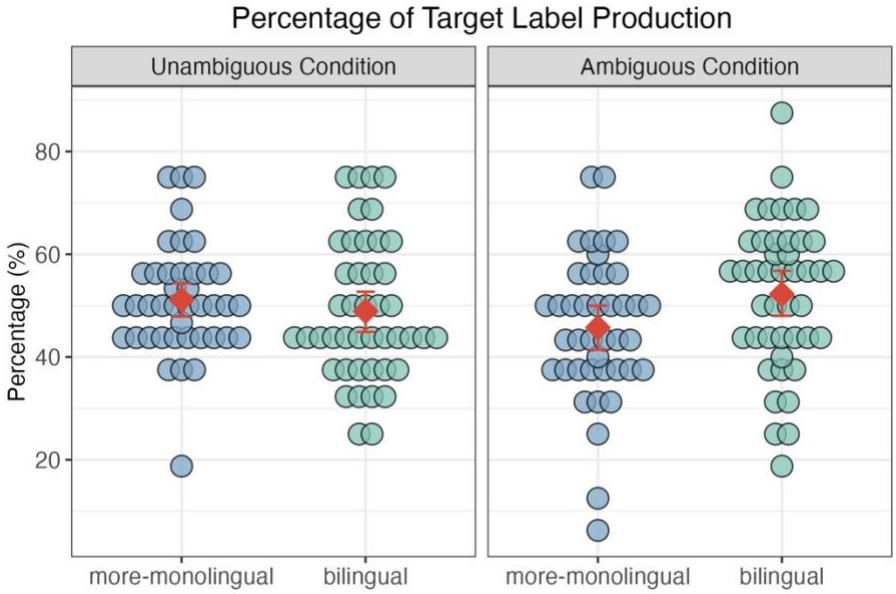

Figure 5: The use of target labels in the unambiguous and ambiguous conditions in Experiment 1. Plotting conventions as in Experiment 1.

Figure 6 shows the percentage of target-label responses across the four ambiguity categories. As in Experiment 1, more-monolinguals used fewest target-label production for homonymy pairs in the ambiguous condition, as indicated by the main effect of Ambiguity Category across the three contrasts (tone-and-segment vs homonymy: b = 1.65, CrI = [0.45, 2.87], pd = 100%; segment-only vs tone-and-segment vs segment-only: b = 0.66, CrI = [-0.42, 1.72], pd = 89%; first-character-only vs segment-only: b = -0.96, CrI = [-2.08, 0.16], pd = 95%). The absence of interactions between condition and ambiguity category as well as between group and ambiguity category suggest that this pattern held across both groups and ambiguity conditions. Crucially, it also indicates that the observed interaction effect between group and condition, i.e. bilinguals produced more target labels than more-monolinguals for ambiguous trials, was broadly consistent across all ambiguity types. Full results of this model are provided in Table 4 in Appendix B.

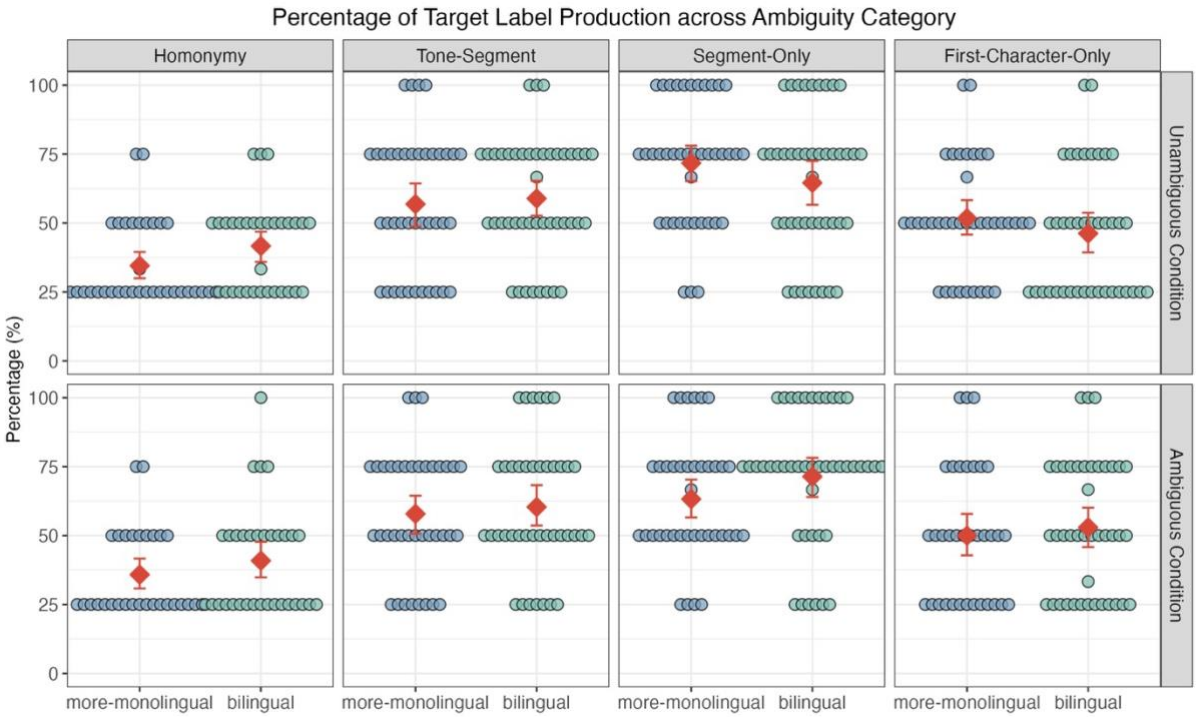

Figure 6: The use of target labels across conditions and ambiguity types in Experiment 2. Plotting conventions as in Experiment 1.

Calibration accuracy

Table 2 summarises the average calibration accuracy (%) for the more-monolingual and bilingual participants who passed both tasks in both calibration sessions. We conducted separate Welch's t-tests for the dot-based task to compare group performance within each session. During the first calibration session, more-monolingual participants showed significantly higher accuracy than bilingual participants ($t(41.60) = 3.17$, $p = .003$). In the second session, more-monolinguals again performed numerically better than bilinguals, but the group difference was only marginally significant ($t(45.44) = 1.91$, $p = 0.06$). While the results of this task may raise some concerns about baseline comparability between groups prior to the main task, this issue is further addressed in the quadrant-based calibration task.

For the quadrant-based task, which shares the same visual layout as the main picture naming task and is thus more directly relevant, we fitted a Bayesian mixed-effects logistic regression to compare performance across sessions. The fixed effects included Calibration Session (first vs second), Group (more-monolingual vs bilingual), and their interactions. Calibration Session was sum-coded (first = -0.5, second = +0.5), and Group was dummy coded with *more-monolingual* as the reference level. The model included by-participant random

intercepts and slopes for Calibration Session, as well as by-item random intercepts and slopes for Calibration Session, Group, and their interaction.

The analysis showed that more-monolinguals did not differ in performance across sessions (b = -0.06, CrI = [-0.44, 0.37], b = 68%). Similarly, there was no reliable difference between the two groups in the second calibration session (b = -0.07, CrI = [-0.57, 0.36], pd = 67%) and no interaction between calibration session and group (b = 0.13, CrI = [-0.18, 0.46], pd = 82%). These results suggest that baseline eye-tracking data quality was comparable between groups at the time of testing. Full results are provided in Table 5 in Appendix B.

Table 2: Calibration Accuracy (Mean %, SD)

Accuracy refers to the proportion of gaze fixations falling within the ROI; SD = standard deviation.

Groups	**First Calibration Session**		**Second Calibration Session**	
	Dot-based	Quadrant-based	Dot-based	Quadrant-based
More-Monolinguals (n = 36)	96.81 (3.82)	87.19 (3.74)	96.31 (4.13)	86.19 (3.90)
Bilinguals (n = 28)	92.28 (6.88)	85.52 (4.92)	93.66 (6.47)	86.17 (4.00)

Saccade behaviour

Figure 7 shows the proportion of saccade in two conditions at the Preview, Pre-Naming, and Post-Naming phases across groups (more-monolinguals vs bilinguals) and verbal responses (non-target labels vs target labels). The upper two panels show the three saccade types in the Ambiguous condition: *target-competitor*, *target-filler*, *competitor-filler*. The lower two panels show the three saccade types in the Unambiguous Condition: *target-filler1*, *target-filler2, filler1-filler2*. As described in Data Analysis, *target-competitor* was the critical saccade in the Ambiguous Condition, whereas *target-filler1* was the critical saccade in the Unambiguous Condition used for comparison.

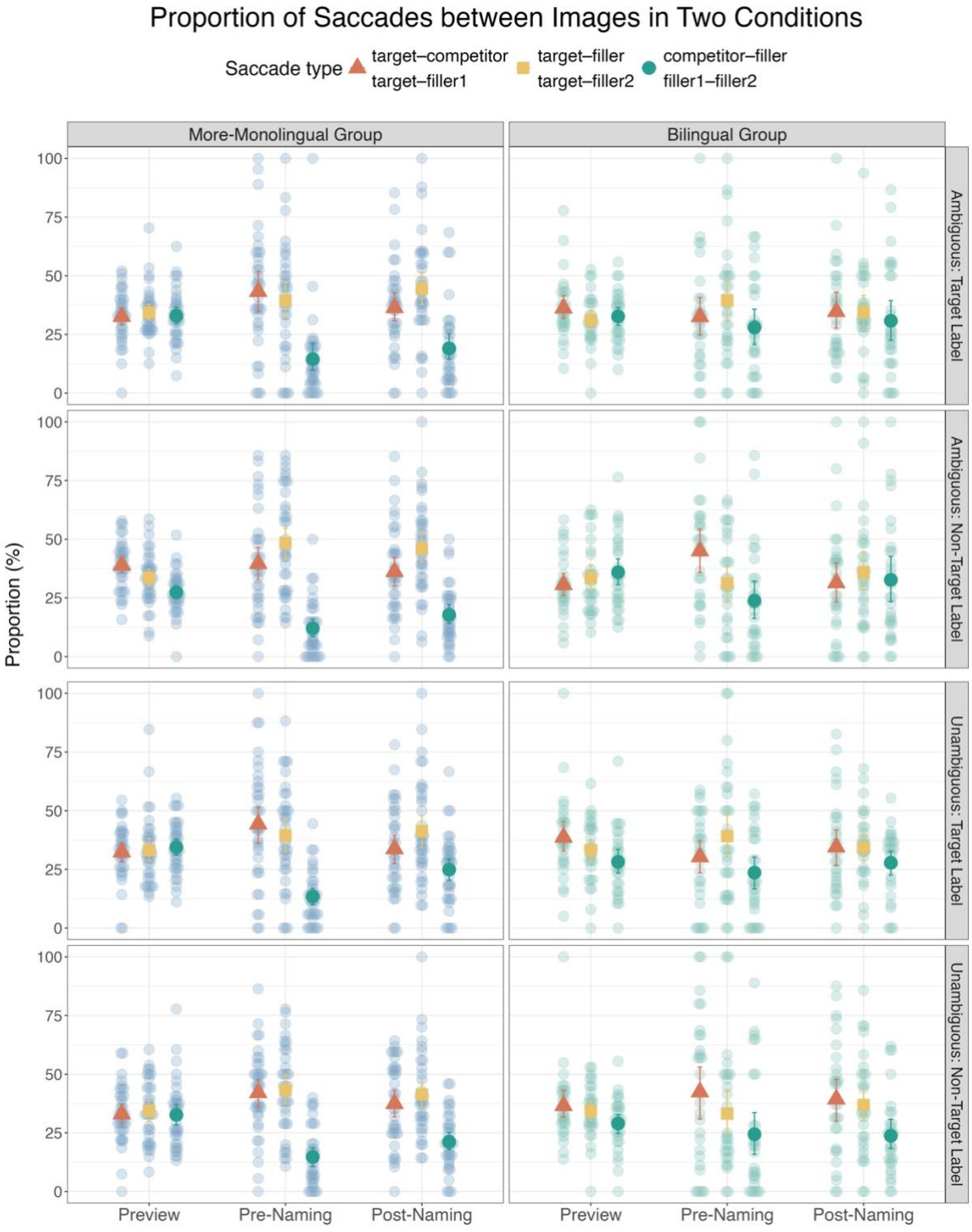

Figure 7: The mean proportion of gaze shift between three image pairs for ambiguous trials (upper two panels, i.e. *target-competitor*, *target-filler*, *competitor-filler*) and for unambiguous trials (lower two panels, i.e. *target-filler1*, *target-filler2*, *filler1-filer2*) across three phases (Preview, Pre-Naming, and Post-Naming), two groups (more-monolingual and bilingual), and two response types (target-label and non-target label) in Experiment 2. Filler and Filler 2 refer to the same image. Each dot represents a speaker's mean proportion of one saccade type. The red triangle, yellow square and green dot represent the grand means for the three saccade types,

respectively, in the two conditions. Error bars indicate bootstrapped 95% confidence intervals of the grand mean.

We fitted a full model that included all four predictors: Phase, Condition, Response Type, and Group, to investigate how these factors interactively shape participants' gaze behaviour. Critical saccades were coded as 1 and all the other saccade types were coded as 0. Response Type and Condition were sum-coded (Response Type: non-target-label = -0.5, target-label = +0.5; Condition: unambiguous = -0.5, ambiguous = +0.5), while the other predictors were dummy coded using *more-monolingual* as reference level for Group and *preview* for Phase. This model included by-participant random intercepts and slopes for Phase, Condition, and their interaction, as well as by-item (i.e. image array) random intercepts and slopes for Phase, Group, and their interaction. Response Type was not included in random effects, as it was examined post-hoc rather than manipulated as part of the experimental design.

Full results are provided in Table 6 in Appendix B. Figure 8 presents the conditional effect of the four predictors on the posterior probability of critical saccades, as estimated by the Bayesian model. Our primary interest was on when (i.e. in which phase) gaze behaviour diverges between ambiguous and unambiguous conditions across groups. Therefore, the results reported here mainly focus on effects that involve Condition, either alone or in interaction with other variables. We first report more-monolingual speakers' gaze patterns across conditions and then compare their behaviours with bilinguals.

At the Preview Phase, more-monolingual speakers tended to make more critical saccades for ambiguous trials than for unambiguous trials, averaging across target and non-target descriptions, although this effect was likely small, since the credible intervals were narrow and close to zero (as indicated by the main effect Condition: b = 0.21, CrI = [0.01, 0.41], pd = 98%). This effect of Condition was further modulated by Response Type: at the Preview Phase, they made fewer saccades for trials where they ultimately used the target descriptions in the Ambiguous Condition, relative to the Unambiguous Condition (as revealed by Condition x Response Type: b = 0.28, CrI = [-0.54, -0.03], pd = 98%). The two-way interaction between Condition and Phase was not reliable at Pre-Naming (b = -0.13, [-0.54, 0.28], pd = 74%) or Post-Naming (b = -0.17, [-0.48, 0.14], pd = 86%) Phases. Similarly, the three-way interaction between Condition, Phase, and Response Type was not robust at later phases (Pre-Naming: b = 0.33, CrI = [-0.14, 0.80], pd = 91%; Post-Naming: b = 0.18, CrI = [-

0.23, 0.57], pd = 81%). These null effects of interaction terms suggest that changes from Preview to later phases and/or across response types did not differ across ambiguity conditions. Overall, more-monolingual participants had more saccade between target and competitor in ambiguous trials, particularly if they avoid making an ambiguous target response, compared to unambiguous trials, suggesting monitoring of/sensitivity to potential ambiguity.

When comparing groups, at the Preview Phase, bilingual speakers made fewer critical saccades than more-monolinguals in the Ambiguous Condition averaging across target and non-target descriptions, as indicated by the two-way interaction between Condition and Group (b = -0.36, CrI = [-0.67, -0.05], pd = 99%). The three-way interaction between Phase, Condition, and Group was not robust at Pre-Naming (b = 0.11, CrI = [-0.61, 0.81], pd = 62%) or Post-Naming (b = 0.06, CrI = [-0.44, 0.55], pd = 59%) Phases, suggesting this effect was roughly constant across phases.

When averaging across Conditions, a different pattern emerged depending on Response Type across groups. The Group x Response Type interaction revealed that at the Preview Phase, bilinguals made *more* saccades than more-monolinguals for trials where they used target-label responses (b = 0.26, CrI = [0.05, 0.47], pd = 99%); but fewer saccades at later phases, as captured by the Phase x Group x Response Type interaction (Pre-Naming: b = -0.60, CrI = [-1.01, -0.19], pd = 100%; Post-Naming: b = -0.41, CrI = [-0.74, -0.09], pd = 99%).

Notably, these effects did not interact reliably with Condition. The three-way interaction involving Condition, Group and Response Type suggests a tendency for bilingual speakers to make more critical saccades on target-label trials in the Ambiguous Condition, relative to the Unambiguous Condition and to more-monolingual speakers; however, the directionality and magnitude of this effect were uncertain, as the credible intervals included zero (b = 0.35, CrI = [-0.04, 0.75], pd = 96%). Similarly, the four-way interaction involving Phase, Condition, Group, and Response Type suggested that bilinguals tended to make fewer critical saccades on target-label production at Pre-Naming and Post-Naming Phases for ambiguous trials; again, these effects were not very robust, as their credible intervals also included zero (Pre-Naming: b = -0.78, CrI = [-1.58, 0.02], pd = 97%; Post-Naming: b = -0.50, CrI = [-1.13, 0.12], pd = 94%).

This suggests that group differences in saccade behaviour depending on responses were broadly consistent across both conditions: bilinguals tended to show more early critical saccades on trials where they ultimately selected (ambiguous) target-label expressions, but reduced critical saccades at later phases, compared to more-monolingual speakers.

In summary, bilinguals and more-monolinguals showed different gaze patterns. At the Preview Phase, bilinguals made fewer critical saccades overall in the Ambiguous Condition than more-monolinguals. This was further modulated by Response Type: bilinguals made more critical saccades than more-monolinguals on trials where they used ambiguous target labels. At later phases, bilinguals decreased critical saccades on those target-label trials more than more-monolinguals. We discuss the implications of their divergent gaze patterns in detail in the Discussion, but our interpretation is that these differences reflect reduced sensitivity to potential ambiguity in our bilingual participants, and instead a facilitative effect whereby the shared label between target and competitor in ambiguous trials in fact facilitates producing their shared (but ambiguous) label for bilinguals.

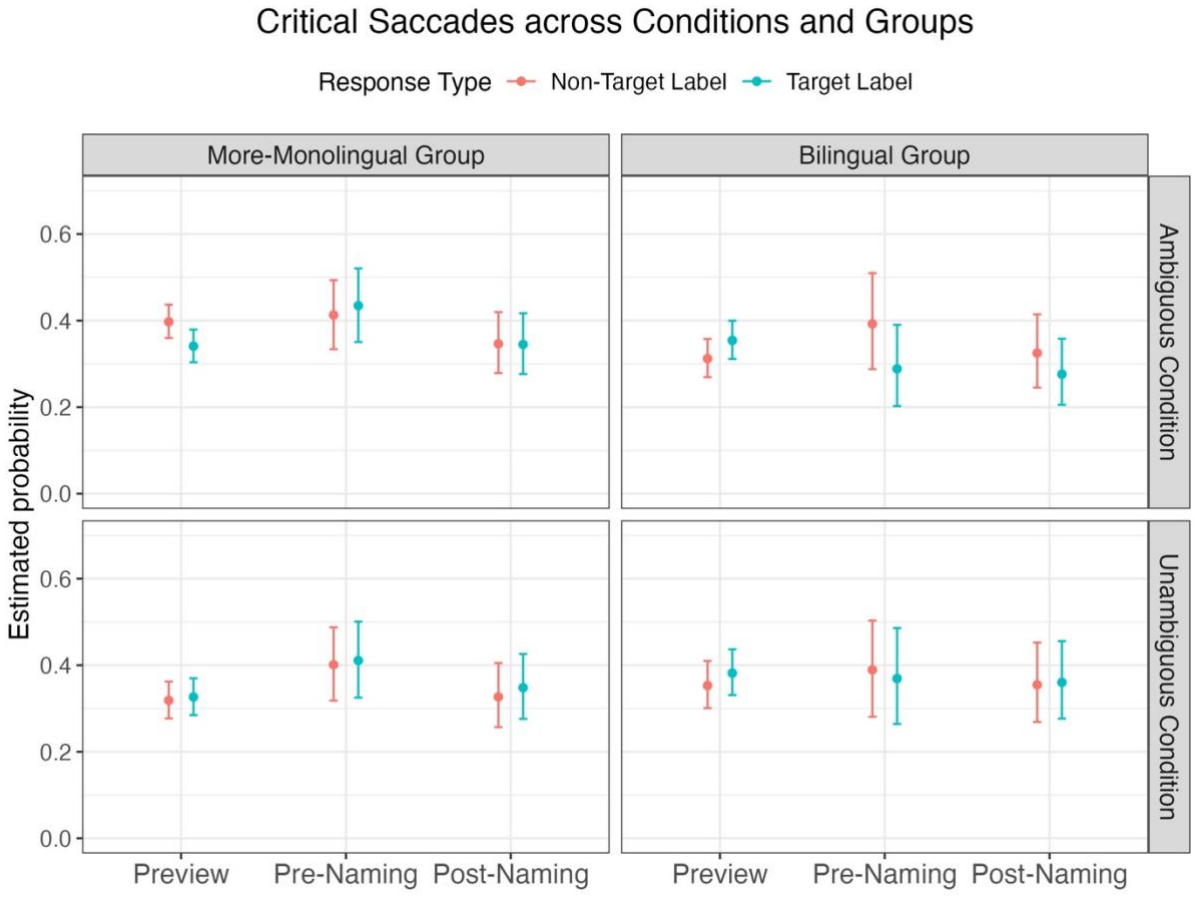

Figure 8: The conditional effects of Phase, Condition, Group, and Response Type on the probability of critical saccades (Ambiguous Condition: *target-competitor*; Unambiguous Condition: *target-filler1*), as estimated by the Bayesian model. The red and blue dots represented estimated mean probability, and error bars indicate the 95% credible intervals.

3.5 Combined Analysis

We ran two additional models combining participants' verbal responses from the two experiments to increase statistical power by maximising our sample size.

In the first model, we included Experiment (Experiment 1 vs Experiment 2), Group and Condition, along with their three-way interaction. In the second model, we added Ambiguity Category and tested the full four-way interaction. Experiment was sum-coded (Experiment 1 = -0.5, Experiment 2 = +0.5), Ambiguity Category was applied successive difference contrast, while the other predictors used default treatment coding, with reference levels set to match those used in Experiment 1 and 2 (i.e. *more-monolingual* for Group, *ambiguous* for Condition).

The results of these two models are presented in Table 7 and 8 in Appendix B. Like the analysis of Experiment 2, the results of the first model again revealed the main effect of Group, suggesting that bilinguals used more target label for ambiguous trials, compared to more-monolinguals (b = 0.38, CrI = [0.06, 0.71], pd = 99%), and a reliable two-way interaction between group and condition, averaging across two experiments (b = 0.36, CrI = [0.05, 0.68], pd = 99%), confirming this tendency for trial ambiguity to affect bilingual and more-monolingual speakers differently. None of the other two-way or three-way interaction terms were credible, suggesting that the change in visual layout between Experiment 1 and 2 did not influence participants' spoken responses.

The results of the second model were also consistent with those of Experiment 2: homonymy pairs elicited fewest target-label production in both conditions, compared to other ambiguity types. The observed interaction effect between group and condition, i.e. bilinguals produced more target labels than more-monolinguals for ambiguous trials, was broadly present across all ambiguity types. Additionally, none of the interaction terms involving Experiment were credible, confirming that the change in visual layout did not affect participants' production.

5.3.6 Discussion

In Experiment 2, we replicated the picture-naming task used in Experiment 1 with a larger sample of L1 Mandarin L2 English bilinguals and their more-monolingual peers and incorporated online eye-tracking to investigate how linguistic ambiguity is processed in real time. Experiment 2 (and the combined analysis of production data from both experiments) revealed a reliable interaction effect between Condition and Group: bilinguals were more likely than more-monolinguals to produce ambiguous labels in the Ambiguous Condition relative to the Unambiguous Condition. This effect seemed to be present broadly for all ambiguity types.

With regard to gaze pattern, bilinguals and more-monolinguals showed different tendencies. During the Preview Phase, compared to more-monolinguals, bilingual participants made fewer critical saccades in the Ambiguous Condition relative to the Unambiguous Condition. This suggests that bilinguals may have engaged in less ambiguity monitoring during early visual processing of the image array. Moreover, their gaze behaviour was response-dependent: bilingual participants were more likely than more-monolinguals to make more critical saccades at Preview on trials where they ultimately produced target expressions. At later phases, they decreased critical saccades on those target-label trials (as revealed by Phase x Group x Response Type).

This gaze pattern suggests that bilinguals may be more likely to direct early attention toward image pairs whose associated labels are more accessible for them (e.g., easier to retrieve and produce), possibly due to their own experience and linguistic accessibility. In the Ambiguous Condition, the label overlap between the target and competitor images likely increased the activation of their shared (ambiguous) label, making it more accessible and thus more likely to be selected. This corresponds to their verbal responses that they were more likely to use the ambiguous label than more-monolinguals in the Ambiguous Condition. At later phases, bilinguals reduced their critical saccades after producing those target expressions, suggesting that once they committed to an effortless, readily accessible expression, they no longer felt the need to further monitor that image pair. Their gaze pattern suggests that bilingual participants may recognise the shared label but appear less motivated to engage in additional monitoring or avoidance of ambiguity.

In contrast, our more-monolingual speakers overall tended to make more critical saccades (i.e. between target and competitor) in the Ambiguous Condition than in the Unambiguous Condition. This suggests their early sensitivity to potential ambiguity, thought the magnitude of that sensitivity can be relatively small (note that Rabagliati and Robertson, 2017, found no reliable difference between ambiguous and unambiguous trials at Preview). Crucially, this early sensitivity was response-dependent. Specifically, for trials where more-monolingual participants ultimately chose a non-target, alternative expression, they were likely to make more critical saccades at Preview, suggesting early sensitivity to ambiguity, compared to trials where they finally chose a default, ambiguous expression. In other words, if they were aware of the potential ambiguity, they were likely to avoid it at naming.

More-monolinguals' gaze pattern did not differ across conditions from Preview to Pre-Naming, nor were they further modulated at Pre-Naming. This absence of interaction effects could be attributed to a levelling effect triggered by cue-driven attention: when the target was indicated, more-monolingual participants may have allocated attention to the target more uniformly across trials. At the Post-Naming Phase, more-monolinguals did not show retrospective monitoring of their speech (i.e. no increase in critical saccades for ambiguous trials), contrary to previous suggestions in Ferreira et al. (2005) and Rabaglliati and Robertson (2017). One possible explanation might be that since our picture naming task was self-paced, more-monolingual participants likely had sufficient time to choose a label that they found satisfactory. On trials where they noticed potential ambiguity early and selected an unambiguous alternative to resolve this ambiguity, their need for visual monitoring at Post-Naming was less strong. Interestingly, on trials where they eventually used a target, ambiguous expression, more-monolinguals did not show increased visual monitoring after naming. As mentioned earlier, more-monolingual speakers may not have initially detected the ambiguity on those trials (as reflected in fewer critical saccades). Once the target was cued, the label shared by the target and competitor may have become activated in their mental representation, leading them to adopt that label without much further re-evaluation.

As noted, bilinguals were more likely than more-monolinguals to adopt those highly activated shared labels for ambiguous trials. However, the cognitive process underlying this choice appeared fundamentally different. Importantly, bilinguals were likely aware of the potential ambiguity initially on those trials where they used ambiguous labels (as revealed by

more critical saccades), but were less inclined to avoid it. Then, why did awareness not lead to avoidance?

Previous research suggests that bilingual experience may lead to increased cognitive demands due to the need to manage two language systems, which often remain co-activated even when one is not currently in use (e.g., Hatzidaki, Branigan, & Pickering, 2010). This constant cross-language activation requires additional control and monitoring, making processing more effortful compared to monolingual language use. Therefore, it makes sense that bilinguals tend to prioritise reducing processing demands when making linguistic choice. One domain in which this becomes particularly relevant is lexical access. Previous research on L1 attrition has shown that changes in one's native language after being immersed in their L2 often emerge most rapidly at the lexical level (e.g. Baus et al., 2013; Jarvis, 2019; Linck et al., 2009; Weltens & Grendel, 1993). One commonly observed phenomenon is the difficulty in retrieving lexical items in their native language that may result from reduced use of their L1 and increased use of their other languages (e.g., Ecke, 2004; Schmid & Jarvis, 2014; Schmid & Yilmaz, 2018). In the context of lexical ambiguity, the presence of a competitor image that shares linguistic features with the target likely activates a shared label, making it more accessible. To avoid such ambiguity, speakers would need to inhibit this activated label and instead retrieve a less active, more specific alternative. This retrieval process becomes more effortful for bilinguals due to lexical access difficulties. As such, it is more cognitively efficient to rely on the shared, ambiguous label that is already activated in the mental lexicon. This may explain why bilingual participants in our study were more likely than more-monolinguals to produce ambiguous expressions under conditions of lexical ambiguity and their gaze patterns suggest early attention to image pairs whose labels are easier for them to retrieve.

These findings further challenge the account by the Pragmatic Principles Violation Hypothesis (Lozano, 2016), which attributes bilinguals' overuse of more explicit referential forms (e.g., overt pronouns) to a communicative preference for clarity over economy. Our current study shows that bilinguals do not always tend to avoid ambiguity more than more-monolinguals. Instead, their behaviour under lexical ambiguity appears to be constrained by cognitive load and lexical accessibility. As discussed earlier, in the context of pronoun usage, overt and null pronouns differ in the strength of cue they provide; overt pronouns help reduce referential uncertainty (even if they do not fully disambiguate in some contexts), whereas

using null pronouns requires speakers to constantly integrate contextual information, placing a heavier burden on them to track referents. In this case, bilinguals may therefore prefer overt pronouns as a way to ease processing demands of tracking referents, rather than as a means to avoid ambiguity.

We conclude by noting some of the limitations and future direction of this study. Like Rabagliati and Robertson (2017), our experiments did not involve genuine conversational interaction since our participants named pictures in isolation rather than communicating with or interpreting another speaker. This controlled setting allowed us to examine cognitive and perceptual processes and revealed different strategies in response to lexical ambiguity among bilinguals and more-monolinguals. However, it leaves open the question of how these patterns would unfold in dialogue when speakers have to adapt to a partner's feedback in real time.

In addition, our use of image stimuli (some with added visual cues, e.g., arrows and circles highlighting features critical to the target label) and varying visual complexity (some fillers are relatively simple easy-to-name images) may have introduced unintended salience effects. As mentioned in Experiment 1, our stimulus creation process began with selecting label pairs from the four ambiguity categories, and then looking for representative images that could elicit those labels. When selecting fillers, we ensured that they did not share phonological or orthographic features with the experimental labels. However, this might have unintentionally led to fillers that were easier to identify or less ambiguous than the experimental items. For instance, in the image array of a homonymy pair, *tai2 tou2* (抬头), meaning either "to lift one's head" or "the header of an invoice", we added visual cues to help clarify the intended meanings. Specifically, in the image illustrating the physical action, an arrow was used to indicate the upward motion of a woman lifting her head; in the image of the invoice, a red circle was used to highlight the header section. While these cues were intended to clarify interpretations, they might also have drawn more attention to the target and competitor images compared to the filler, which was semantically unrelated but left unmarked. Nonetheless, these design features alone are unlikely to account for the gaze patterns that varied across conditions and response types among more-monolinguals and bilinguals. If participants' attention was driven by these static visual features, we would expect similar patterns across both groups. Although the observed differences in gaze patterns depending on

responses and between groups are unlikely to result solely from such low-level visual features, future work should take perceptual properties into consideration and ensure that all images in a trial are matched for visual salience.

Another limitation is that participants in the lexical ambiguity task were not tested on reference production in anaphoric contexts (i.e. the choice of null pronouns, overt pronouns, and NPs). As a result, it remains unclear whether bilingual individuals who overuse overt pronouns and/or NPs also tend to rely on ambiguous labels in contexts of lexical ambiguity. Future within-subject designs are needed to determine whether these two tendencies would co-occur within individuals, which would offer stronger evidence for the underlying mechanisms that drive bilinguals' referential behaviour.

Despite these limitations, our study points to several conclusions about how native Mandarin speakers handle lexical ambiguity in spoken Mandarin and how bilinguals differ from more-monolinguals in this task in both production and online processing. As in prior work, we found that speakers often do not proactively detect linguistic ambiguities. However, our more-monolingual participants did engage in proactive monitoring: they identified potential ambiguity early when they later chose an unambiguous alternative. This effect was reliable, though uncertain in magnitude. In contrast, they did not show reliable retrospective monitoring after speaking when they ultimately used the ambiguous label. By contrast, our L1 Mandarin L2 English bilinguals do not tend to monitor and avoid such ambiguities and were more likely than more-monolinguals to produce ambiguous expressions. Their gaze pattern and linguistic choices suggest a higher tolerance for ambiguity, possibly driven by processing efficiency considerations.

Taken together, these findings do not support the hypothesis that bilinguals would rather than be redundant than risk being ambiguous in general; instead, they support the hypothesis that bilinguals' linguistic choices are likely driven by a desire to minimize processing effort. Our study also sheds light on the mechanisms of lexical attrition in one's native language as a result of learning a second language, providing additional evidence of the processing difficulties bilinguals face when accessing L1 lexicon. These findings not only underscore how reduced L1 use and L2 immersion can reshape real-time processing of the L1 lexicon, but also demonstrate how bilinguals adapt to evolving language experience.

4 Conclusion

We conducted two experiments to understand the underlying mechanism for bilinguals' preference, seen in other work, for more explicit forms, by investigating how L1 Mandarin L2 English speakers manage lexical ambiguity in spoken Mandarin, a subtle aspect of referential domain. We tested two hypotheses: (1) bilinguals would rather be redundant than ambiguous or (2) they use more explicit signals to reduce processing load. Our findings support the second hypothesis. Specifically, bilinguals tended to use more ambiguous labels in ambiguous trials, compared to more-monolingual speakers, and participants eye-tracking behaviour further reveals that bilinguals tended to direct early attention to image pairs whose shared labels are more accessible to them. In contrast, more-monolinguals showed proactive monitoring of ambiguity, particularly when they avoided making an ambiguous response.

References

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2017). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9

Argyri, E., & Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition, 10*(1), 79-99. https://doi.org/10.1017/S1366728906002835

Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of Pragmatics, 16*(5), 443-463. https://doi.org/10.1016/0378-2166(91)90136-L

Baus, C., Costa, A., & Carreiras. (2013). On the effects of second language immersion on first language production. *Acta Psychologica. 142*(3), 402-409. https://doi.org/10.1016/j.actpsy.2013.01.010

Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language & Linguistic Theory, 25*(4), 657–689. https://doi.org/10.1007/s11049-007-9026-9

Blackwell, S. E. (1998). Constraints on Spanish NP Anaphora: The Syntactic versus the Pragmatic Domain. *Hispania, 81*(3), 606–618. https://doi.org/10.2307/345683

Boersma, P., & Weenink, D. (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.23, http://www.praat.org/

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1). https://doi.org/10.18637/jss.v080.i01

De Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an Open-Source CollaborativeEcosystem of Behavioral Experiments. *Journal of Open Source Software*, *8*(85), 5351. https://doi.org/10.21105/joss.05351

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition, 20*(5), 917-930. https://doi.org/10.1017/S1366728916000547

Ecke, P. (2004). Language attrition and theories of forgetting: A cross-disciplinary review. *International Journal of Bilingualism*, *8*(3), 321–354. https://doi.org/10.1177/13670069040080030901

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, 54, 554-573.

https://doi.org/10.1016/j.jml.2005.12.009

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition, 96*, 263–284. https://doi.org/10.1016/j.cognition.2004.09.002

Geluykens, R. (2013). *Pragmatics of Discourse Anaphora in English: Evidence from Conversational Repair*. Walter de Gruyter.

Gürel, A. (2019). Null and Overt Pronouns in Language Attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 250–263). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.21

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds) *Studies in Syntax and Semantics III: Speech Acts* (pp. 183-98). Academic Press.

Jarvis, A. (2019). Lexical attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 250–263). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.20

Kaltsa, M., Tsimpli, I. M., & Rothman, J. (2015). Exploring the source of differences and similarities in L1 attrition and heritage speaker competence: Evidence from pronominal resolution. *Lingua, 164*, 266–288. https://doi.org/10.1016/j.lingua.2015.06.002

Levinson, S. C. (1987a). Minimization and conversational inference. In In M. Bertuccelli Papi, & J. Verschueren (Eds), *The Pragmatic Perspective: Selected papers from the 1985 International Pragmatics Conference* (pp. 61-129). John Benjamins.

Levinson, S. C. (1987b). Pragmatics and the Grammar of Anaphora: A Partial Pragmatic Reduction of Binding and Control Phenomena. *Journal of Linguistics, 23*, 379- 434. https://www.jstor.org/stable/4175896

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12), 1507-1515. https://doi.org/10.1111/j.1467-9280.2009.02480.x

Li, X. J., Li, X. S., Qu, Q. (2022). Predicting phonology in language comprehension: Evidence from the visual world eye-tracking task in Mandarin Chinese. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(5), 531-547. https://doi.org/10.1037/xhp0000999

Liang, X., Huang, F., Liu, D., & Xu, M. (2024). Brain representations of lexical ambiguity: Disentangling homonymy, polysemy, and their meanings. *Brain and Language, 253,* Article 105426. https://doi.org/10.1016/j.bandl.2024.105426

Liu, Y., Sorace, A., & Smith, K. (2025). Beyond Crosslinguistic Influence: Mandarin Speakers with Exposure to Null-subject Languages Nonetheless Use Fewer Null Pronouns in Mandarin. *Proceedings of the 47th Annual Meeting of the Cognitive Science Society*, 878-884. https://escholarship.org/uc/item/4h69k203

Lozano, C. (2016). Pragmatic principles in anaphora resolution at the syntax-discourse interface: advanced English learners of Spanish in the CEDEL2 corpus. In M. Alonso Ramos (Ed.), *Spanish learner corpus Research: Current trends and future perspectives* (pp. 236-265). John Benjamins. https://doi.org/10.1075/scl.78.09loz

Lozano, C. (2018). The Development of Anaphora Resolution at the Syntax-Discourse Interface: Pronominal Subjects in Greek Learners of Spanish. *Journal of Psycholinguist Research, 47*, 411–430. https://doi.org/10.1007/s10936-017-9541-8

Lozano, C., & Quesada, T. (2023). What corpus data reveal about the Position of Antecedent Strategy: Anaphora resolution in Spanish monolinguals and L1 English-L2 Spanish bilinguals. *Frontiers in Psychology, 14*, 1246710. https://doi.org/10.3389/fpsyg.2023.1246710

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software, 4*(40), 1541. https://doi.org/10.21105/joss.01541

Margaza, P., & Bel, A. (2006). Null Subjects at the Syntax-Pragmatics Interface: Evidence from Spanish Interlanguage of Greek Speakers. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.*), Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference* (pp. 88–97). Cascadilla Press. https://www.lingref.com/cpp/gasla/8/paper1491.pdf

Martin-Villena, F. (2023). L1 morphosyntactic attrition at the early stages: evidence from production, interpretation, and processing of subject referring expressions in L1 Spanish-L2 English instructed and immersed bilinguals. Doctoral dissertation, Universidad De Granada. https://digibug.ugr.es/handle/10481/81920

Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition, 7*(2), 125–142. https://doi.org/10.1017/S1366728904001464

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, Jeff., & Hays, J. (2016).

WebGazer: Scalable Webcam Eye Tracking Using User Interactions. *Proceedings of The 25th International Joint Conference on Artificial Intelligence,* 3839-3845. https://cs.brown.edu/people/apapouts/papers/ijcai2016webgazer.pdf

Qu, Q., Li, X., Wei, W. (2024, September 5-7). Dissociating the pre-activation of orthography and phonology during reading: Evidence from EEG representational similarity analysis. [Conference presentation]. The 30th Architectures and Mechanisms for Language Processing, Edinburgh, the United Kingdom. https://virtual.oxfordabstracts.com/event/31397/submission/145

R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing [Computer software]. https://www.r-project.org/

Rabagliati, H., & Robertson, A (2017). How do children learn to avoid referential ambiguity? Insights from eye-tracking. *Journal of Memory and Language*. 94, 15-27. http://dx.doi.org/10.1016/j.jml.2016.09.007

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficient-based analysis of color overspecification. *Frontiers in Psychology*, *7*, 153. https://doi.org/10.3389/fpsyg.2016.00153

Serratrice, L., Sorace, A., & Paoli, S. (2004). Crosslinguistic influence at the syntax–pragmatics interface: Subjects and objects in English–Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition*, *7*(3), 183–205. https://doi.org/10.1017/S1366728904001610

Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, *17*(4), 729–748. https://doi.org/10.1017/S1366728913000771

Schmid, M. S., & Yılmaz, G. (2018). Predictors of Language Dominance: An Integrated Analysis of First Language Attrition and Second Language Acquisition in Late Bilinguals. *Frontiers in Psychology*, *9*, 1306. https://doi.org/10.3389/fpsyg.2018.01306

Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech, 58*(2), 131–151. https://doi.org/10.1177/0023830914522956

Slim, M. S., & Hartsuiker, R. J. (2023). Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIbex and WebGazer.js. *Behavior Research Methods, 55*, 3786-3804. https://link.springer.com/article/10.3758/s13428-022-01989-z

Shin, N.L. & Smith Cairns, H. (2009). Subject pronouns in child Spanish and continuity of reference. In Collentine, J., M. García, B. Lafford & F. Marcos Marín (Eds), *Selected Proceedings of the 11th Hispanic Linguistics Symposium* (pp. 155–164). Cascadilla Press.

Solomyak, O., & Marantz, A. (2009). Lexical access in early stages of visual word processing: A sing-trial correlational MEG study of heteronym recognition. *Brain and Language, 108*(3), 191-196. https://doi.org/10.1016/j.bandl.2008.09.004

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism, 1*(1), 1–33. https://doi.org/10.1075/lab.1.1.01sor

Sorace, A. (2016). Referring expressions and executive functions in bilingualism. *Linguistic Approaches to Bilingualism, 6*(5), 669–684. https://doi.org/10.1075/lab.15055.sor

Sorace, A. (2019). L1 attrition in a wider perspective. *Second Language Research, 36*(2), pp. 203-206. https://doi.org/10.1177/0267658319895571

Sorace, A., Serratrice, L., Filiaci, F., & Baldo, M. (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua, 119*(3), 460–477. https://doi.org/10.1016/j.lingua.2008.09.008

Tal, S., Smith, K., Arnon, I., & Culbertson, J. (2023). Communicative efficiency is present in young children and becomes more adult-like with age. *Proceedings of the Annual Meeting of the Cognitive Science Society, 45*(45), 1446-1452. https://escholarship.org/uc/item/7mm0z6fk

Trott, S., & Bergen, B. (2020). Why do human languages have homophones? *Cognition, 205*, Article 104449. https://doi.org/10.1016/j.cognition.2020.104449

Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. (2004). First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, *8*(3), 257–277. https://doi.org/10.1177/13670069040080030601

Weltens, B., & Grendel, M. (1993). Attrition of vocabulary knowledge. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 135–156). John Benjamins. https://doi.org/10.1075/sibil.6.08wel

Wu, S.-L. (2020). Crosslinguistic influence in development of reference realization: A comparison of foreign language learners and heritage language learners. *Chinese as a Second Language Research*, *9*(2), 227–257. https://doi.org/10.1515/caslar-2020-2009

Zhang, A., & Kwon, N. (2022). The interpretational preferences of null and overt pronouns in Chinese. *Journal of Linguistics*, *58*(3), 649–676.

https://doi.org/10.1017/S0022226721000402

Appendix A

Experiment 1: Questionnaire Response

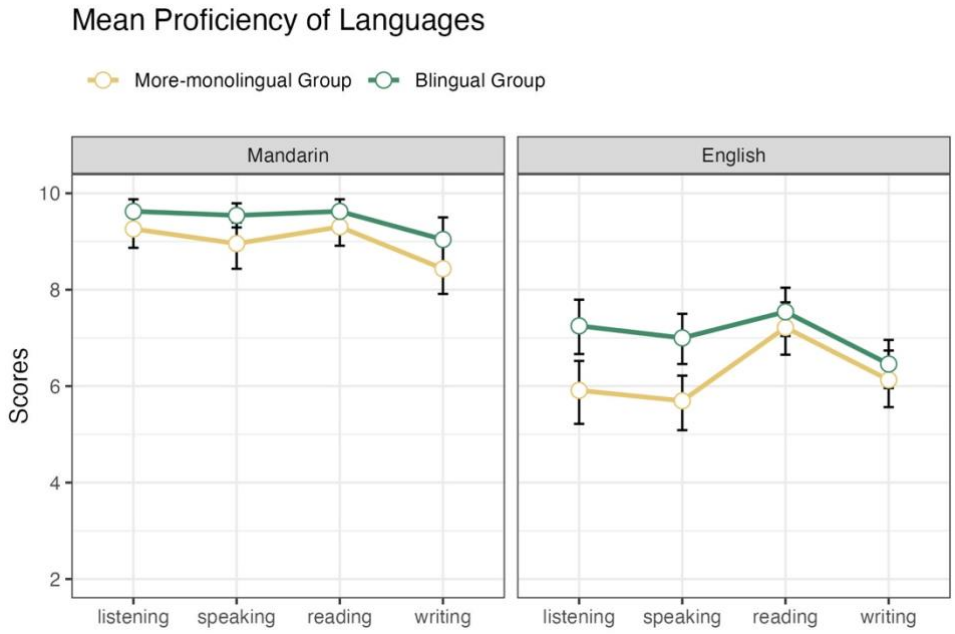

Figure 1: The mean proficiency scores in Mandarin and English in four skills (listening, speaking, reading, and writing) across the more-monolingual and bilingual groups in Experiment 1. Error bars indicate bootstrapped 95% confidence intervals of the mean.

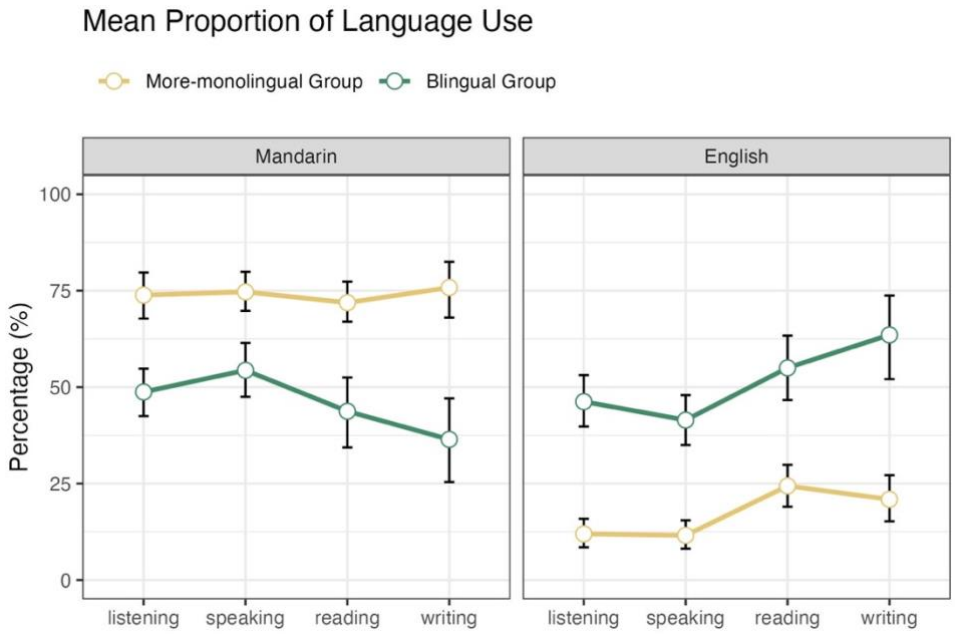

Figure 2: The mean percentage of language use in Mandarin and English in the respective four skills across groups in Experiment 1. Plotting conventions as in Figure 1.

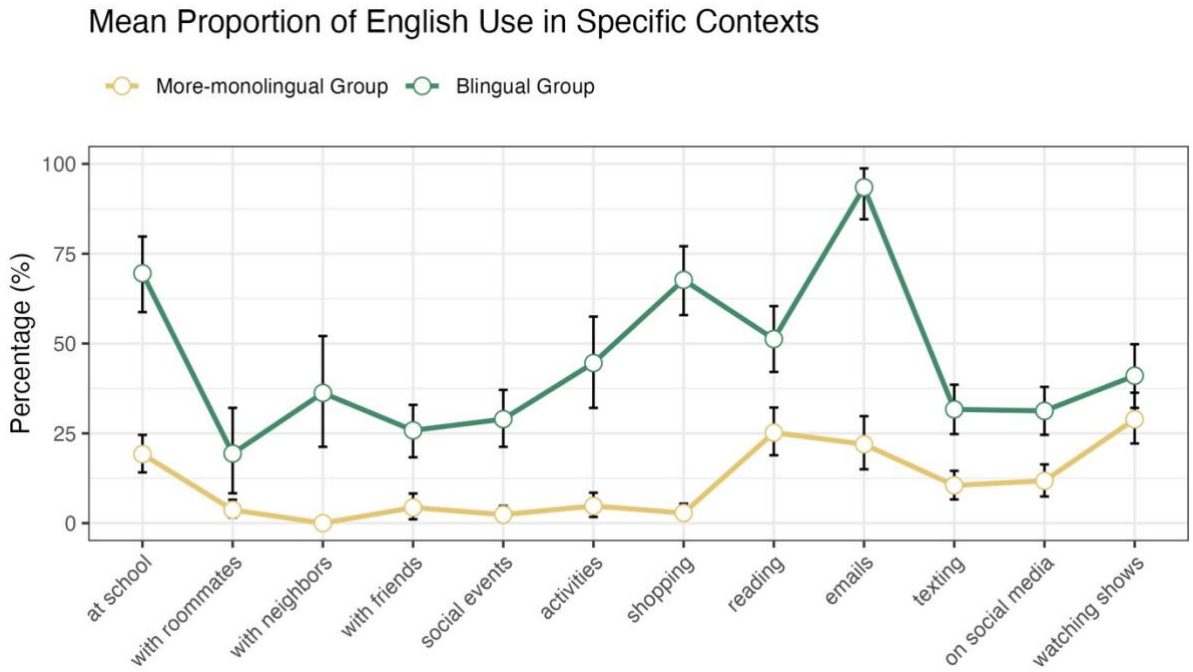

Figure 3: The mean percentage of language use in English in 12 specific daily situations across groups in Experiment 1. Plotting conventions as in Figure 1 and 2.

Experiment 2: Questionnaire Response

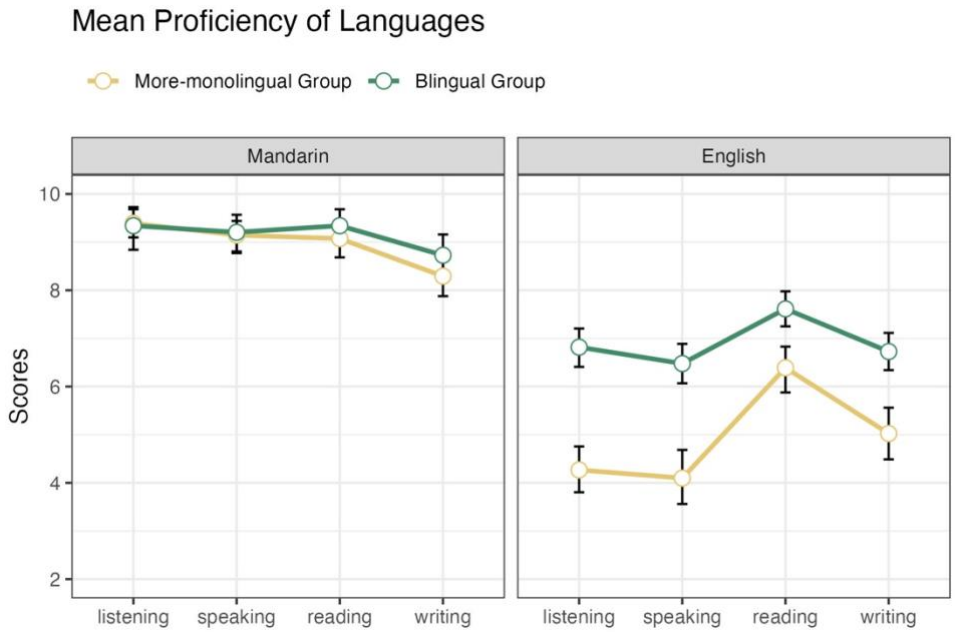

Figure 4: The mean proficiency scores in Mandarin and English in four skills across groups in Experiment 2. Plotting conventions as in previous figures.

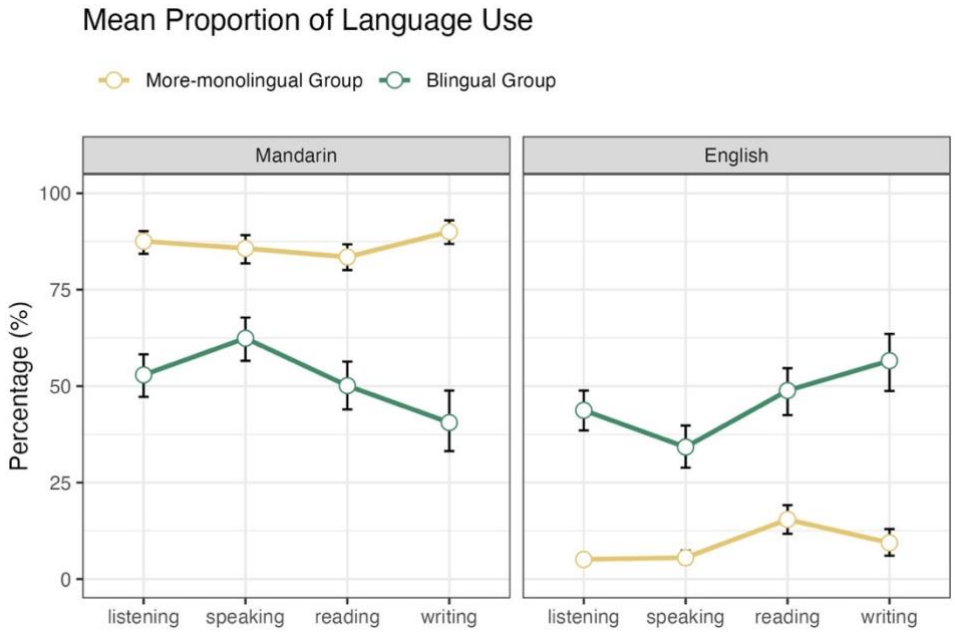

Figure 5: The mean percentage of language use in Mandarin and English in the respective four skills across two groups in Experiment 2. Plotting conventions as in previous figures.

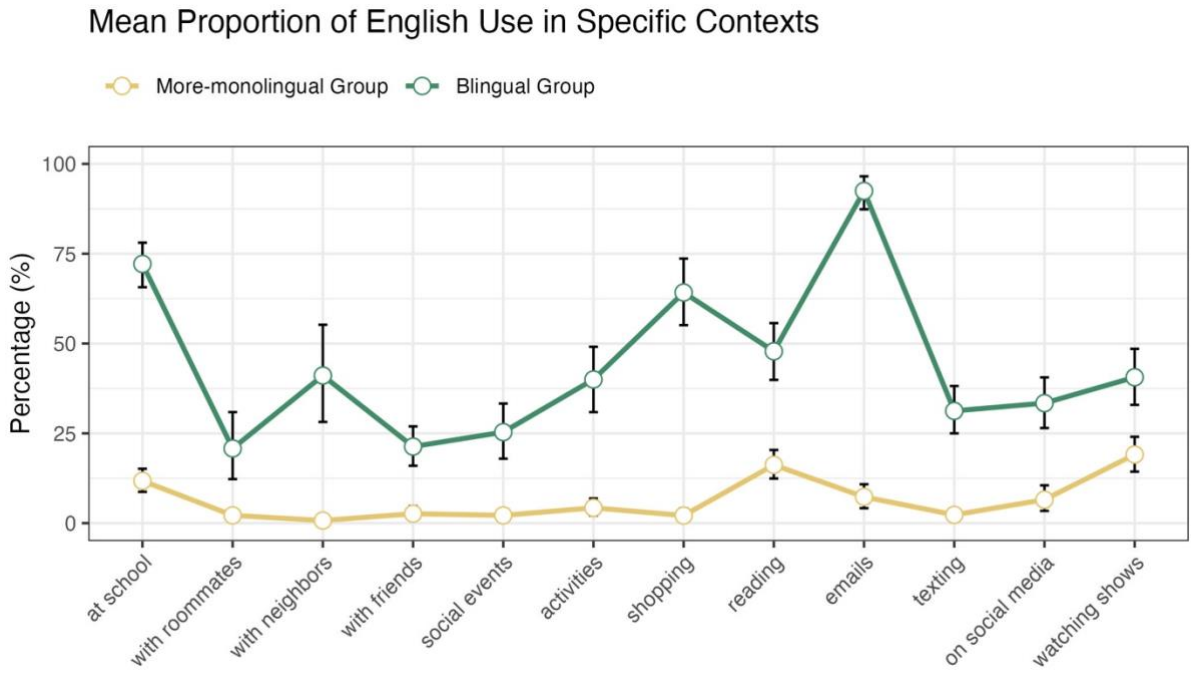

Figure 6: The mean percentage of language use in English in 12 specific daily situations in Experiment 2. Plotting conventions as in previous figures.

Appendix B

Experiment 1: Verbal Responses

Model: Response Type ~ Condition * Group + (1 + Condition | Participant) + (1 + Group | Item)

Table 1: The outputs of the Bayesian logistic model for effects of Condition and Group on Response Type in Experiment 1. Positive values indicate an increased usage of target labels, whereas negative values indicate a decreased usage of target labels. Condition = Ambiguous and Group = More-monolingual were set as reference levels, respectively.

Predictors	**Estimates**	**95% CrI**	**PD (%)**
Intercept (Ambiguous, More-monolingual)	0.05	[-0.52, 0.61]	57
Condition = Unambiguous vs Ambiguous	0.16	[-0.56, 0.88]	67
Group = Bilingual vs More-monolingual	0.25	[-0.28, 0.79]	82
Condition * Group = Ambiguous **:** Bilingual	-0.18	[-0.75, 0.39]	74

Model: Response Type ~ Condition * Group * Ambiguity Category + (1 + Condition * Ambiguity Category | Participant) + (1 + Group | Item)

Table 2: The outputs of the Bayesian logistic model for effects of Condition, Group, and Ambiguity Category on Response Type in Experiment 1. Positive values indicate an increased usage of target labels, whereas negative values indicate a decreased usage of target labels. Condition = Ambiguous and Group = More-monolingual were set as reference levels, respectively. Ambiguity Category was applied with successive difference contrast.

Predictors	**Estimates**	**95% CrI**	**PD (%)**
Intercept (Ambiguous, More-monolingual, Averaging across Ambiguity Category)	-0.08	[-0.64, 0.48]	61
Condition = Ambiguous vs Unambiguous	0.11	[-0.59, 0.81]	62
Group = Bilingual vs More-monolingual	0.26	[-0.31, 0.83]	82
Ambiguity Category = Tone-and-Segment vs Homonymy	1.52	[0.33, 2.76]	99
Ambiguity Category = Segment-Only vs Tone-and-Segment	0.92	[-0.13, 2.02]	96

Ambiguity Category = First-Character-Only vs Segment-Only	-0.97	[-2.06, 0.12]	96
Condition * Group = Unambiguous **:** Bilingual	-0.19	[-0.79, 0.41]	74
Condition * Ambiguity Category = Unambiguous **:** Tone-and-Segment vs Homonymy	-0.17	[-1.74, 1.42]	58
Condition * Ambiguity Category = Unambiguous **:** Segment-Only vs Tone-and-Segment	0.38	[-1.08, 1.82]	70
Condition * Ambiguity Category = Unambiguous **:** Segment-Only vs Tone-and-Segment	-0.53	[-2.03, 0.92]	75
Group * Ambiguity Category = Bilingual **:** Tone-and-Segment vs Homonymy	0.11	[-0.91, 1.15]	59
Group * Ambiguity Category = Bilingual **:** Segment-Only vs Tone-and-Segment	-0.93	[-1.90, 0.06]	97
Group * Ambiguity Category = Bilingual **:** First-Character-Only vs Segment-Only	0.75	[-0.27, 1.77]	92
Condition * Group * Ambiguity Category = Unambiguous **:** Bilingual **:** Tone-and-Segment vs Homonymy	0.54	[-0.89, 1.97]	78
Condition * Group * Ambiguity Category = Unambiguous **:** Bilingual **:** Segment-Only vs Tone-and-Segment	0.19	[-1.18, 1.54]	61
Condition * Group * Ambiguity Category = Unambiguous **:** Bilingual **:** First-Character-Only vs Segment-Only	0.02	[-1.34, 1.36]	51

Experiment 2: Verbal Responses

Model: Response Type ~ Condition * Group + (1 + Condition | Participant) + (1 + Group | Item)

Table 3: The outputs of the Bayesian logistic model for effects of Condition and Group on Response Type in Experiment 2. Positive values indicate an increased usage of overlapping labels, whereas negative values indicate a decreased usage of overlapping labels. Condition = Ambiguous and Group = More-monolingual were set as reference levels, respectively.

Predictors	**Estimates**	**95% CrI**	**PD (%)**
Intercept (Ambiguous, More-monolingual)	-0.04	[-0.58, 0.51]	56
Condition = Unambiguous vs Ambiguous	0.23	[-0.47, 0.97]	73
Group = Bilingual vs More-monolingual	0.44	[0.06, 0.82]	99
Condition * Group = Unambiguous : Bilingual	**-0.47**	**[-0.88, -0.05]**	**99**

Model: Response Type ~ Condition * Group * Ambiguity Category + (1 + Condition * Ambiguity Category | Participant) + (1 + Group | Item)

Table 4: The outputs of the Bayesian logistic model for effects of Condition, Group, and Ambiguity Category on Response Type in Experiment 2. Positive values indicate an increased usage of target labels, whereas negative values indicate a decreased usage of target labels. **Condition = Ambiguous and** Group = More-monolingual were set as reference levels, respectively. Ambiguity Category was applied with successive difference contrast.

Predictors	**Estimates**	**95% CrI**	**PD (%)**
Intercept (Ambiguous, More-monolingual, Averaging across Ambiguity Category)	-0.24	[-0.76, 0.27]	82
Condition = Ambiguous vs Unambiguous	0.23	[-0.45, 0.91]	75
Group = Bilingual vs More-monolingual	**0.46**	**[0.09, 0.84]**	**99**
Ambiguity Category = Tone-and-Segment vs Homonymy	**1.65**	**[0.45, 2.87]**	**100**
Ambiguity Category = Segment-Only vs Tone-and-Segment	0.66	[-0.42, 1.72]	89
Ambiguity Category = First-Character-Only vs Segment-Only	-0.96	[-2.08, 0.16]	95

Condition * Group = Unambiguous : Bilingual	**-0.51**	**[-0.92, -0.09]**	**99**
Condition * Ambiguity Category = Unambiguous **:** Tone-and-Segment vs Homonymy	0.06	[-1.54, 1.68]	52
Condition * Ambiguity Category = Unambiguous **:** Segment-Only vs Tone-and-Segment	0.43	[-1.02, 1.88]	72
Condition * Ambiguity Category = Unambiguous **:** Segment-Only vs Tone-and-Segment	-0.32	[-1.82, 1.18]	66
Group * Ambiguity Category = Bilingual **:** Tone-and-Segment vs Homonymy	-0.42	[-1.18, 0.35]	86
Group * Ambiguity Category = Bilingual **:** Segment-Only vs Tone-and-Segment	0.37	[-0.37, 1.14]	83
Group * Ambiguity Category = Bilingual **:** First-Character-Only vs Segment-Only	-0.12	[-0.89, 0.66]	62
Condition * Group * Ambiguity Category = Unambiguous **:** Bilingual **:** Tone-and-Segment vs Homonymy	0.78	[-0.27, 1.84]	93
Condition * Group * Ambiguity Category = Unambiguous : Bilingual : Segment-Only vs Tone-and-Segment	**-1.08**	**[-2.14, -0.02]**	**98**
Condition * Group * Ambiguity Category = Unambiguous **:** Bilingual **:** First-Character-Only vs Segment-Only	0.16	[-0.89, 1.22]	62

Experiment 2: Eye-tracking data

Quadrant-based Calibration

Model: Accuracy ~ Session * Group + (1 + Session | Participan | Participant) + (1 + Session * Group | Item)

Table 5: The outputs of the Bayesian logistic model for effects of Session and Group on Accuracy in Experiment 2. Positive values indicate increased calibration accuracy, whereas negative values indicate decreased calibration accuracy. Session was sum-coded (First = -0.5, Second = +0.5). Group was dummy-coded with More-monolingual set as the reference level.

Predictors	**Estimates**	**95% CrI**	**PD (%)**
Intercept (More-monolingual, averaging across Session)	**1.92**	**[1.66, 2.15]**	**100**
Session = First vs Second	-0.06	[-0.44, 0.37]	68
Group = Bilingual vs More-monolingual	-0.07	[-0.57, 0.36]	67
Session * Group = First : Bilingual	0.13	[-0.18, 0.46]	82

Picture Naming Saccades

Model: Critical Saccades ~ Phase * Condition * Group * Response Type + (1 + Phase * Condition | Participant) + (1 + Phase * Group | Item)

Table 6: The outputs of the Bayesian logistic model for effects of Phase (Preview, Pre-Naming, Naming, Post-Naming), Condition (Unambiguous, Ambiguous), Group (More-monolingual, Bilingual), and Response Type (Target Label vs Non-Target Label) on Critical Saccades in Experiment 2. Positive values indicate an increase in critical saccades (Ambiguous: target-competitor; Unambiguous: target-filler1), whereas negative values indicate a decrease in critical saccades. Phase = Preview and Group = More-monolingual were set as reference levels, respectively. Response Type was sum-coded (Target Label = +0.5, Non-Target Label = -0.5). Condition was sum coded (Ambiguous = +0.5, Unambiguous = -0.5).

Predictors	**Estimates**	**95% CrI**	**PD (%)**
Intercept (Preview, More-monolingual, Averaged across Response Types and Conditions)	**-0.64**	**[-0.76, -0.52]**	**100**
Phase = Pre-Naming vs Preview	**0.29**	**[0.12, 0.47]**	**100**
Phase = Post-Naming vs Preview	-0.02	[-0.22, 0.18]	57
Condition = Ambiguous vs Unambiguous	**0.21**	**[0.01, 0.41]**	**98**

Group = Bilingual vs More-monolingual	0.02	[-0.15, 0.20]	59
Response Type = Target vs Non-Target Label	-0.10	[-0.23, 0.03]	94
Phase * Condition = Pre-Naming : Ambiguous	-0.13	[-0.54, 0.29]	73
Phase * Condition = Post-Naming : Ambiguous	-0.16	[-0.47, 0.14]	85
Phase * Group = Pre-Naming : Bilingual	-0.26	[-0.58, 0.07]	94
Phase * Group = Post-Naming : Bilingual	-0.08	[-0.39, 0.23]	70
Condition * Group = Ambiguous : Bilingual	**-0.36**	**[-0.67, -0.05]**	**99**
Phase * Response Type = Pre-Naming : Target Label	0.17	[-0.08, 0.41]	91
Phase * Response Type = Post-Naming : Target Label	0.15	[-0.06, 0.35]	92
Condition * Response Type = Ambiguous : Target Label	**-0.28**	**[-0.53, -0.02]**	**98**
Group * Response Type = Bilingual : Target Label	**0.26**	**[0.06, 0.46]**	**99**
Phase * Condition * Group = Pre-Naming **:** **A**mbiguous **:** Bilingual	0.10	[-0.62, 0.82]	61
Phase * Condition * Group = Post-Naming **:** **A**mbiguous **:** Bilingual	0.05	[-0.44, 0.54]	58
Phase * Condition * Response Type = Pre-Naming **:** **A**mbiguous **:** Target Label	0.33	[-0.15, 0.80]	91
Phase * Condition * Response Type = Post-Naming **:** **A**mbiguous **:** Target Label	0.17	[-0.23, 0.57]	80
Phase * Group * Response Type = Pre-Naming : Bilingual : Target Label	**-0.59**	**[-1.01, -0.18]**	**100**
Phase * Group * Response Type = Post-Naming : Bilingual : Target Label	**-0.41**	**[-0.74, -0.08]**	**99**
Condition * Group * Response Type = Ambiguous **:** Bilingual **:** Target Label	0.35	[-0.04, 0.74]	96
Phase * Condition * Group * Response Type = Pre-Naming **:** **A**mbiguous **:** Bilingual **:** Target Label	-0.78	[-1.60, 0.02]	97
Phase * Condition * Group * Response Type = Post-Naming **:** **A**mbiguous **:** Bilingual **:** Target Label	-0.50	[-1.13, 0.14]	94

Combined Experiment 1 and 2: Verbal Responses

Model: Response Type ~ Condition * Group * Experiment + (1 + Condition | Participant) + (1 + Group * Experiment | Item)

Table 7: The outputs of the Bayesian logistic model for effects of Condition, Group, and Experiment on Response Type in combined analysis. Positive values indicate an increased usage of overlapping labels, whereas negative values indicate a decreased usage of overlapping labels. **Condition = Ambiguous** and Group = More-monolingual were set as reference levels, respectively. Experiment was sum-coded (Experiment 1 = -0.5, Experiment 2 = +0.5).

Predictors	**Estimates**	**95% CrI**	**PD (%)**
Intercept (Ambiguous, More-monolingual, Averaged across Experiments)	0.01	[-0.48, 0.49]	52
Condition = Unambiguous vs Ambiguous	0.19	[-0.46, 0.85]	72
Group = Bilingual vs More-monolingual	**0.38**	**[0.06, 0.71]**	**99**
Experiment = Experiment 2 vs Experiment 1	-0.09	[-0.55, 0.37]	65
Condition * Group = Unambiguous : Bilingual	**-0.36**	**[0.71, -0.02]**	**98**
Condition * Experiment = Unambiguous **:** Experiment 2	0.12	[-0.39, 0.61]	68
Group * Experiment = Bilingual **:** Experiment 2	0.10	[-0.52, 0.72]	63
Condition * Group * Experiment = Unambiguous **:** Bilingual **:** Experiment 2	-0.16	[-0.83, 0.51]	69

Model: Response Type ~ Condition * Group * Ambiguity Category * Experiment + (1 + Condition * Ambiguity Category | Participant) + (1 + Group * Experiment | Item)

Table 8: The outputs of the Bayesian logistic model for effects of Condition, Group, and Experiment on Response Type in combined analysis. Positive values indicate an increased usage of overlapping labels, whereas negative values indicate a decreased usage of overlapping labels. Condition = Ambiguous and Group = More-monolingual were set as reference levels, respectively. Ambiguity Category was applied with successive difference contrast. Experiment was sum-coded (Experiment 1 = -0.5, Experiment 2 = +0.5).

Predictors	**Estimates**	**95% CrI**	**PD (%)**
Intercept (Ambiguous, More-monolingual, Averaged across Experiments and Ambiguity Category)	-0.18	[-0.65, 0.29]	78
Condition = Unambiguous vs Ambiguous	0.19	[-0.44, 0.81]	72

Group = Bilingual vs More-monolingual	**0.40**	**[0.08, 0.71]**	**99**
Ambiguity Category = Tone-and-Segment vs Homonymy	**1.61**	**[0.49, 2.75]**	**100**
Ambiguity Category = Segment-Only vs Tone-and-Segment	0.80	[-0.20, 1.80]	94
Ambiguity Category = First-Character-Only vs Segment-Only	-0.91	[-1.94, 0.14]	96
Experiment = Experiment 2 vs Experiment 1	-0.13	[-0.58, 0.33]	70
Condition * Group = Unambiguous : Bilingual	**-0.39**	**[-0.74, -0.04]**	**99**
Condition * Ambiguity Category = Unambiguous **:** Tone-and-Segment vs Homonymy	-0.13	[-1.61, 1.37]	57
Condition * Ambiguity Category = Unambiguous **:** Segment-Only vs Tone-and-Segment	0.47	[-0.87, 1.84]	75
Condition * Ambiguity Category = Unambiguous **:** First-Character-Only vs Segment-Only	-0.40	[-1.84, 1.00]	71
Group * Ambiguity Category = Bilingual **:** Tone-and-Segment vs Homonymy	-0.24	[-0.90, 0.41]	76
Group * Ambiguity Category = Bilingual **:** Segment-Only vs Tone-and-Segment	-0.19	[-0.83, 0.45]	72
Group * Ambiguity Category = Bilingual **:** First-Character-Only vs Segment-Only	0.30	[-0.36, 0.96]	82
Condition * Experiment = Unambiguous **:** Experiment 2	0.13	[-0.37, 0.64]	68
Group * Experiment = Bilingual **:** Experiment 2	0.12	[-0.49, 0.73]	65
Ambiguity Category * Experiment = Tone-and-Segment vs Homonymy **:** Experiment 2	0.28	[-0.61, 1.17]	74
Ambiguity Category * Experiment = Segment-Only vs Tone-and-Segment **:** Experiment 2	-0.23	[-1.08, 0.66]	70
Ambiguity Category * Experiment = First-Character-Only vs Segment-Only **:** Experiment 2	-0.16	[-1.04, 0.73]	64
Condition * Group * Ambiguity Category = Unambiguous **:** Bilingual **:** Tone-and-Segment vs Homonymy	0.77	[-0.14, 1.69]	95

Condition * Group * Ambiguity Category = Unambiguous : Bilingual : Segment-Only vs Tone-and-Segment	-0.58	[-1.46, 0.31]	90
Condition * Group * Ambiguity Category = Unambiguous : Bilingual : First-Character-Only vs Segment-Only	0.19	[-0.70, 1.10]	66
Condition * Group * Experiment = Unambiguous : Bilingual : Experiment 2	-0.18	[-0.87, 0.47]	70
Condition * Ambiguity Category * Experiment = Unambiguous : Tone-and-Segment vs Homonymy : Experiment 2	0.24	[-0.97, 1.44]	65
Condition * Ambiguity Category * Experiment = Unambiguous : Segment-Only vs Tone-and-Segment : Experiment 2	-0.02	[-1.18, 1.15]	51
Condition * Ambiguity Category * Experiment = Unambiguous : First-Character-Only vs Segment-Only : Experiment 2	0.23	[-0.99, 1.40]	66
Group * Ambiguity Category * Experiment = Bilingual : Tone-and-Segment vs Homonymy : Experiment 2	-0.41	[-1.54, 0.71]	76
Group * Ambiguity Category * Experiment = Bilingual : Segment-Only vs Tone-and-Segment : Experiment 2	1.00	[-0.10, 2.09]	96
Group * Ambiguity Category * Experiment = Bilingual : First-Character-Only vs Segment-Only : Experiment 2	-0.65	[-1.78, 0.44]	87
Condition * Group * Ambiguity Category * Experiment = Unambiguous : Bilingual : Tone-and-segment vs Homonymy : Experiment 2	0.13	[-1.31, 1.63]	57
Condition * Group * Ambiguity Category * Experiment = Unambiguous : Bilingual : Segment-Only vs Tone-and-segment : Experiment 2	-0.85	[-2.26, 0.57]	88

Condition * Group * Ambiguity Category * Experiment = Unambiguous **:** Bilingual **:** First-Character-Only vs Segment-Only **:** Experiment 2	-0.15	[-1.60, 1.33]	58