

Running title: Individual binding preferences in adolescent HSs

Full title: To bind or not to bind: Individual differences in pronominal processing among adolescent Mandarin-English heritage speakers

Authors: Jiuzhou HAO^{1,2,3}, Vincent DELUCA², Jason ROTHMAN^{2, 3, 4}

¹School of Philosophy, Psychology and Language Sciences, University of Edinburgh, United Kingdom

²Department of Language and Culture (ISK), UiT The Arctic University of Norway, Norway

³Department of Linguistics and English Language, Lancaster University, United Kingdom

⁴Nebrija Research Centre in Cognition, Nebrija University, Spain

Acknowledgements. We thank the participants who made this research possible. Our special thanks go to the enthusiastic families who advertised the study on our behalf.

Funding statement: This project was funded by the European Union's Horizon Europe research and innovation program under the Marie Skłodowska-Curie grant agreement No 101104834 and the Trond Mohn Foundation, under the Center for Language Brain and Learning (C-LaBL) grant No. TMS2023UiT01. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Address for correspondence: Jiuzhou Hao, jhao@ed.ac.uk, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, United Kingdom

Competing interests: The author(s) declare none.

Disclosure of use of AI tools: None declared.

Keywords: Individual differences; binding principles; adolescent heritage speakers, simplex and complex anaphors, pronominals

Introduction

To date, heritage language (HL) acquisition research has primarily focused on either early childhood development or adult competence/performance outcomes (Kupisch & Rothman, 2018; Montrul, 2016; Polinsky, 2018). As a result, HL development and use in adolescence remain understudied, despite compelling reasons to the contrary (Bayram et al., 2021; Korkus & Vihman, 2024; Minkov et al., 2019). Indeed, having a clear picture of HL development over late childhood and adolescence is essential for capturing the full developmental trajectory of HL bilingualism, in particular for linguistic domains where sub-properties are generally known to be gradually acquired in stages over time (e.g., properties related to pronouns). The fact that adolescent HL bilinguals are differentially subject to (drastic) changes in their input and usage patterns, emerging sociolinguistic realities with increased personal agency as well as maturation in other cognitive and psychological abilities results in a rather unique natural laboratory whereby studying their development over this transitional period (from childhood to adulthood) can be informative in multifarious ways.

Heritage speakers (HSs) are early bilinguals who acquire a HL from birth with naturalistic exposure at home or their immediate communities. Importantly, the HL is not the language of the larger society (Montrul, 2016; Rothman, 2009). At the macro-group level, a substantial body of research has documented that adult HSs often perform differently compared to first language (L1)-dominant speakers of the same language (Montrul, 2016; Polinsky, 2018). To investigate the source of these differences, researchers have proposed a comparative developmental approach, examining child HSs and adult HSs relative to monolingual “baselines” (Montrul & Polinsky, 2021; Polinsky, 2018; Polinsky & Scontras, 2020). The logic behind this approach is that if both child and adult HSs pattern similarly and differently from monolinguals at the same time, the difference reflects a differential acquisition trajectory and outcomes between HSs and L1-dominant users. In contrast, if child HSs are more similar to monolingual children but differ from adult HSs, then HL attrition may explain the adult HSs’ outcomes (Polinsky, 2018).

While this comparative approach has yielded important insights, it implicitly assumes a linear or categorical developmental path and often (unwittingly) treats HS populations as internally homogeneous. However, an accumulating body of research highlights that HSs exhibit striking individual differences (IDs) to a degree that is rarely observed among neurotypical L1-dominant users or even L2 populations (De Houwer, 2023; Rothman et al., 2023). These differences span multiple linguistic domains (Daskalaki et al., 2019) and are shaped by a complex interaction of child-internal factors, e.g., cognitive abilities, and child-

external factors, e.g., input quantity and quality (see Paradis, 2023 for a summary). For the comparative framework to be truly explanatory, it must be embedded within a broader developmental model that not only traces language outcomes across the lifespan but also accounts for the heterogeneity observed within HS groups.

Under the comparative approach, another recurring question in the HL bilingualism literature concerns whether all linguistic domains are equally vulnerable to input reduction (Polinsky & Scontras, 2020). One influential proposal is that linguistic structures situated at the interface of syntax with other domains, especially discourse-pragmatics, often the same ones fully converged upon later in L1-dominant child acquisition, are more susceptible to reduced input than properties governed strictly by narrow syntax. This idea is formalized in the Interface Hypothesis (IH), which maintains that morphosyntactic properties requiring integration with discourse-pragmatic information are particularly vulnerable in bilingual grammars (Sorace, 2011), mostly likely for processing related reasons. In addition, structures involving long-distance dependencies, such as object-relative clauses and *wh*-movement, are thought to impose greater difficulties for HSs, a phenomenon sometimes referred to as “the distance problem” (Polinsky & Scontras, 2020). This distance problem suggests that the integration of elements across larger syntactic spans may interact with the vulnerability of interface structures, creating a cumulative challenge for bilingual grammars. The IH and the distance problem have been empirically tested in both child and adult HSs and other bilinguals, but findings supporting its predictions remain mixed (Daskalaki et al., 2019; Hao, Chondrogianni, et al., 2024; Leal et al., 2014).

These mixed results may not be surprising when considering that the comparative approach tends to focus on group-level patterns while often overlooking the substantial variability within HS populations. By averaging across individuals, this approach may obscure meaningful variability in how different structures are processed and acquired by different individuals. An ID approach offers a powerful complementary lens to refine and test the IH and the distance problem. From the ID perspective, narrow syntactic structures and local dependencies are expected to be more uniform across HSs, whereas interface structures and long-distance dependencies are more susceptible to variation.

Additionally, while there has been an upsurge in research adopting an ID approach in the HL literature, doing so in combination with online processing methods trails significantly behind still. Online processing methods measure how language users respond to linguistic information in real time, offering insights into automatic, time-sensitive mechanisms that underpin language comprehension/use. In contrast, offline methods, such as grammaticality

judgment tasks or elicited production tasks, are more likely to be influenced by language users' metalinguistic awareness, task strategies, and explicit knowledge. This distinction is particularly relevant for HSs, whose performance is known to be highly sensitive to such factors (Polinsky, 2018). Online methods therefore offer a more direct window into HL knowledge and are essential for evaluating how ID factors shape language comprehension and use beyond what offline tasks can reveal (Bayram et al., 2021).

The current study investigates how IDs in cognitive and input-related factors affect online HL processing in adolescent Mandarin-English heritage speakers. Using a visual world eye-tracking paradigm, we focus on three types of Mandarin pronouns: pronominals (*ta* "he/she/it", a long-distance dependency located at the syntax-pragmatics interface), simplex reflexives (*ziji* "self", a long-distance dependency located at the syntax-semantics-pragmatic interface), and complex reflexives (*taziji* "himself/herself/itself", a local dependency governed by narrow syntax). By examining how adolescent HSs interpret these forms in real time, we evaluate the extent to which working memory, inhibition, and HL exposure and use patterns modulate HL processing. In doing so, we aim to contribute to a more developmentally grounded, cognitively informed model of HL bilingualism.

ID Factors Modulating HL Development and Processing

Individual differences (IDs) factors can be broadly categorized into user-internal and user-external dimensions (see Paradis, 2023 for a summary). Internal factors refer to the cognitive capacities that language users bring to the task of language acquisition and use, such as working memory, inhibitory control, and other executive functions. These factors influence how efficiently learners can process, store, and retrieve linguistic information. In contrast, external factors encompass the broader socioenvironmental context of language exposure and use. These include proximal factors, such as the amount and quality of direct input in the HL and opportunities for its use, as well as distal factors related to the larger environment, such as socioeconomic status, which can shape the availability and richness of proximal experiences. It is worth noting that while ID factors should, in principle, affect language learners and users more generally, their effects on HL processing need not reflect the same underlying mechanisms across populations. HSs occupy a unique sociolinguistic and developmental position, characterized by early bilingual exposure, long-term dominance shifts, and often sustained reduction in naturalistic HL input. As a result, ID effects in this population are likely to arise from interactions between user-internal cognitive resources and user-external experiential factors that are not directly comparable to those in other bilingual or monolingual

populations. For example, an effect of working memory in HL processing, where input is reduced, variable, and largely naturalistic, may reflect different underlying mechanisms from working-memory effects observed in L1-dominant speakers with full input or in L2 learners whose reduced input is often instruction-driven (see also Cunnings, 2017, on the differential role of working memory in L1-dominant and L2 users). To avoid generalizations that might incorrectly imply shared underlying mechanisms, we therefore limit our discussion of ID effects to HSs.

Starting with (language) user-internal factors, a growing body of research has demonstrated that working memory (WM) plays a positive role in bilinguals' language abilities across a wide range of linguistic domains (e.g., vocabulary, morphosyntax, and narrative skills) and across task modalities (i.e., offline comprehension, production, and online measures). However, the role of WM in HL development and use has received comparatively less attention. Among the few existing studies, Paradis et al. (2020) found that HL vocabulary size in child HSs was positively associated with WM capacity, while Soto-Corominas et al. (2021) reported a similar positive relationship between WM and HL sentence repetition accuracy across a range of morphosyntactic structures. These findings suggest that WM supports HL development at least in young children. While there is good reason to hypothesize a strong relationship between HL processing and WM capacity for particular domains of grammar, namely those that tax memory such as, for example, lexical retrieval, long distance dependencies, those that require integration between grammar and discourse, etc., the relationship between WM and HL processing is severely understudied, and thus unclear. In terms of grammar proper, Bice and Kroll (2021) stand out as a singular study directly examining the role of WM on HL processing. Their findings showed that WM capacity correlated with sensitivity to subject-verb agreement violations in L1-dominant users, but not in adult HSs. While this stands out in contrast to Paradis et al. (2020) and Soto-Corominas et al. (2021), the discrepancy may be task-specific, structure-specific, or age-specific. This highlights the need for carefully considering the linguistic domains and developmental stages under investigation.

Another user-internal factor that is particularly relevant to the current study is inhibitory control, or inhibition. While the role of inhibition in bilingualism has been extensively studied in the context of language switching and lexical selection as well as in research on domain-general cognitive advantages associated with bilingual experience, relatively little attention has been paid to how inhibition supports real-time sentence processing, particularly in HSs. Yet, examining inhibition at the processing level holds significant promise for addressing a central question in bilingualism: how does cross-linguistic influence manifest during language

comprehension, and what cognitive mechanisms help bilinguals manage competing representations from their two languages?

At the processing level, cross-linguistic influence has been reported in terms of HSs' use of the majority language processing strategies even when processing the HL (see Chondrogianni, 2023 for a summary). In such cases, HSs may rely on cues or parsing strategies that are more consistent with their societal language than with those typically employed by dominant speakers of their HL. Several theoretical models of bilingual sentence processing account for this phenomenon by emphasizing the role of cue-based transfer. For instance, both the cue-based retrieval model (e.g., Cunnings, 2017) and the Unified Competition Model (e.g., MacWhinney, 2018) posit that bilinguals interpret sentences by drawing on cues associated with both of their languages. When the two languages differ in cues, bilinguals may transfer cue preferences from one language to the other, resulting in non-target-like processing. Therefore, when HSs process HL structures that diverge from the majority language, those with stronger inhibitory control may show greater apparent success in suppressing default strategies that align with the societal dominant language.

Turning to user-external factors, a substantial body of evidence highlights the importance of HL input quantity (a proximal factor typically measured as current or cumulative exposure to and use of the HL) in shaping HL development and use, particularly as reflected in offline measures across a wide range of linguistic domains (Chondrogianni & Daskalaki, 2023; Daskalaki et al., 2019; Kubota et al., 2025; Paradis et al., 2020; Soto-Corominas et al., 2021). However, the role of HL input quantity on HL processing remains less clear. The emerging literature presents mixed findings, suggesting that the role of HL input quantity may vary across developmental stages or types of tasks. For example, Hao, Chondrogianni et al. (2024) found that among pre-teenage Mandarin-English HSs, HL input quantity significantly predicted performance on offline comprehension and production of various non-canonical structures. However, input quantity did not modulate online processing of the same structures in a self-paced listening with picture verification task. In contrast, studies with adult HSs have reported more consistent effects of input. For instance, using the visual world eye-tracking paradigm, Hao, Kubota, et al. (2024) found that input modulated the processing strategy preferences (see also Karaca et al., 2024). Moreover, emerging neuroimaging work (EEG), also shows a positive correlation between HL grammatical processing and quantitative as well as qualitative aspects of HL input (Hao et al., 2025).

These divergent findings suggest that the role of HL input in processing may be developmentally mediated, and that adolescence could represent a transitional phase in which

the relationship between input and real-time processing begins to emerge or reorganize. Moreover, compared to self-paced listening with picture verification tasks, the visual world eye-tracking and EEG paradigms, at least as implemented in the above studies, can offer a more naturalistic and certainly more temporally fine-grained measure of language processing. They track participants' real-time attention to referents as they listen to/read (spoken) language, allowing the detection of subtle effects of input on processing. Additionally, unlike self-paced listening with picture verification tasks, eye-tracking and EEG do not require an additional metalinguistic or verification task at the end of each trial. This minimizes task-related demands and reduces the likelihood that processing is influenced by response strategies, affective factors (argued to be particularly relevant for HS, Polinsky, 2018) or metalinguistic knowledge, making it particularly suitable for capturing the automatic aspects of HL processing.

The present study focuses on three language user-level factors: WM, inhibition, and HL input quantity. While other language user-level factors may also contribute to HL processing individual differences, including additional predictors would require a substantially larger sample size to ensure adequate statistical power and avoid overfitting. This is especially true as we are also interested in how IDs potentially differentially manifest themselves across different linguistic domains. Moreover, many user-external factors, such as input quantity and input richness, tend to be highly correlated, complicating model specification and interpretability.

To maintain analytical clarity while achieving a robust understanding of the role of WM, inhibition, and HL input quantity, the present study controls for other user-level factors to the extent possible. As detailed in the Participants section, we accounted for HL education, parental language background, and socioeconomic status (SES), three distal factors that have been shown to influence HL development and use (see Paradis, 2023 for a review). By focusing on theoretically motivated core predictors while minimizing potential confounds, this study aims to provide a precise and developmentally sensitive account of how internal and external factors shape real-time HL processing in adolescence.

Pronoun Systems in English and Mandarin

PRONOUNS are linguistic devices whose referential interpretation depends on other elements within the sentence and/or discourse. That is, their meanings are not fixed but are BOUND to and/or CO-REFERENCED with other noun phrases or entities. Pronouns include REFLEXIVES (sometimes referred to as ANAPHORS), such as *himself*, and PRONOMINALS (sometimes referred to reductively as PRONOUNS) such as *he*. Reflexives and pronominals have different

distributions. For example, while the reflexive *himself* in (1) must be interpreted as *Joe* but not *Jack*, the pronominal *him* in (2) must not refer to *Joe* but can be interpreted as *Jack* or some other third person male in the discourse.

(1) Jack thinks Joe likes himself.

(2) Jack thinks Joe likes him.

Chomsky (1993) proposed the BINDING PRINCIPLE A AND B to capture the distribution of reflexives and pronominals respectively. More specifically, Binding Principle A suggests that a reflexive is bound in its governing category/a reflexive is locally bound (the minimal category that contains the reflexive, assigns a thematic role or syntactic Case to the reflexive, and is a SUBJECT). Binding Principle B stipulates that a pronominal must be free within its governing category. While in both (1) and (2), *Jack* and *Joe* C-COMMAND *himself* and *him*, *Joe* is realized within the governing category of the pronouns, but *Jack* is not. In other terms, *Joe* is a local antecedent of the pronouns and *Jack* is a long-distance (LD) antecedent and pronominals can bind with a LD antecedent but cannot bind with a local antecedent and reflexives must bind with a local antecedent.

While Binding Principle A reliably accounts for the distribution of reflexives in ARGUMENT positions in English, Binding Principle B appears to face some challenges, i.e., pronominals governed by Principle B often exhibit behavior that suggests greater interpretive flexibility than Binding Principle B alone predicts. Consider example (3):

(3) I know what Jack and Joe have in common. Jack adores him, and Joe adores him too.

In this case, the second pronominal *him* for many can refer to *Joe*, its apparent local antecedent, seemingly violating Principle B (as a bound variable interpretation). However, such examples are often judged acceptable in context, especially when contrastive focus or parallel discourse structure is present. This example highlights the difference between binding and co-reference (see Reuland, 2006 for a summary). Binding involves a syntactic dependency between a pronoun and its antecedent, constrained by structural rules such as locality and c-command. In contrast, co-reference refers to cases where two expressions refer to the same entity but are not syntactically dependent on each other. Building on this distinction, Grodzinsky and Reinhart (1993) proposed the Rule I, which formalizes a principle of interpretive economy: when both a bound variable reading (via syntactic binding) and a coreferential reading (via discourse) are available and yield the same interpretation, the coreferential reading is blocked. In example (2), *him* cannot be bound by *Joe* (as it violates Principle B) nor be co-referenced with *Joe* (this co-reference reading is identical to a bound reading, violating Rule I). In example (3), however, while the binding reading is ruled out by

Principle B, the co-referential reading (A and B like B) is not blocked by Rule I as it differs from the bound variable reading (A likes A and B likes B), allowing a local co-referential reading. These patterns demonstrate that (long-distance) pronominal resolution draws on both syntactic constraints and pragmatic computations: while syntax constrains the space of possible antecedents, pragmatic and discourse principles determine which referent is ultimately selected.

Under such an account, reflexives, governed by Principle A, are interpreted via syntactic binding: once the structural dependency is established, the reflexive's reference is determined compositionally. In contrast, pronominals are referentially independent and interpreted through discourse co-reference (Rule I) once syntactic constraints (Principle B) have ruled out illicit binding. This asymmetry explains why reflexives are handled within narrow syntax, whereas pronominals require computation at the syntax–pragmatics interface.

Mandarin has three forms of pronouns, the pronominal *ta* “he/she/it” (PRO henceforth), the complex reflexive *taziji* “himself/herself/itself” (SELF henceforth) and the simplex reflexive *ziji* “self” (SE henceforth). The pronominal *ta*, like its English counterpart *him*, is governed by Binding Principle B and Rule I (which blocks co-reference between *ta* and its local antecedent when a complex reflexive *taziji* yields the same interpretation). The complex reflexive *taziji*, corresponding closely to English *himself*, is subject to Binding Principle A, requiring local syntactic binding. In contrast, the SE, *ziji*, lacking in English, can refer either to a LD antecedent or to a local antecedent, violating Binding Principle A. The status of the SE in Mandarin remains highly debated without theoretical consensus.

Some approaches treat the SE as a syntactic anaphor in both LD and local readings (see Cole et al., 2006 for a summary). The basic idea is that LD reading of the SE is a result of Logical Form (LF)-movement where the SE is moved to the T(ense) position at LF, where it receives its features from the subject through Specifier-Head agreement. Under this approach, even LD binding is ultimately local, conforming to the Binding Principle A. However, this movement-based approach fails to account for some empirical observations, such as island constraints and blocking effects (see Wang & Pan, 2021 for a summary of criticisms of the movement approach). One influential alternative approach to the SE proposes that the SE functions as a syntactic anaphor following Binding Principle A and giving rise to local binding and as a pragmatic logophor leading to a LD reading (C. T. J. Huang & Liu, 2000). Logophors impose a consciousness requirement requiring the antecedent being conscious of the relevant event being reported. For example, in example (4), the LD antecedent *Jack* must be aware of the claim made in the embedded clause as it reflects his own view, i.e., the consciousness requirement is met, allowing the SE to take on the LD antecedent as its referent. Under this

non-uniform approach, the interpretation of the LD SE is influenced by syntactic structure, verb semantics (e.g., whether the verb is logophoric), and discourse-level factors (e.g., perspective alignment).

In the present study, as we are interested in IDs in modulating the processing of structures at the interface vs. at the narrow syntax, all main verbs in the experiment are logophoric, leading to a preference for LD binding. Thus, the LD interpretation of SE arises at the syntax–semantics–pragmatics interface: syntax constrains possible antecedents, verb semantics encode logophoricity, and discourse/pragmatic context ultimately selects the referent.

- (4) 杰克 认为 乔 喜欢 自己。
Jack renwei Joe xihuan ziji
Jack believe Joe likes SELF
Jack believes that Joe likes him/himself.

The Acquisition and Processing of English and Mandarin Pronouns

In terms of the acquisition of pronouns in L1-dominant children, both English- and Mandarin-speaking children have been shown to have acquired Binding Principle A from a very young age (around 5 y.o.). That is, they correctly interpret reflexives to a local c-commanding antecedent (Chien, 1992; Chien & Wexler, 1990; Clackson et al., 2011). In contrast, at the same age, English-speaking children have been consistently shown to erroneously accept local bindings of pronominals, a phenomena often referred to as the Delay of Principle B Effect (see Thornton & Wexler, 1999 for a review). This has been attributed either to children’s immature development of pragmatic principles governing pronominal resolution (e.g., Rule I), despite having acquired Binding Principle B (e.g., Chien & Wexler, 1990), or to domain-general limitations such as working memory constraints (e.g., Kim & Yoon, 2020). Studies testing the Delay of Principle B Effect in Mandarin, however, has received mixed results (Chien & Lust, 2006). It is typically not until the age of 9 that children begin to show adult-like performance in correctly rejecting local bindings of pronominals.

As for the development of the Mandarin SE, while children before the age of 4 show unsystematic performance patterns, children from the age of 5 predominantly co-index the SE with local antecedents than with LD antecedents (Chien & Lust, 2006). This local preference is also found in adults. However, individual studies show high degree of variations, e.g., the mean acceptance rates of LD readings range from under 40% to over 90% and even the mean acceptance rates of local readings can range from under 70% to 90% (see Chen & Ionin, 2023

for a summary). Such inter-study variation could be attributed to the different methodologies used in these studies, e.g., truth value judgement task vs. picture-biasing sentence acceptability judgement task, different verbs (logophoric vs. generic verbs), etc., to potential individual variation in executive function (particularly WM) and/or the importance placed on syntax vs. discourse factors (see Kim & Yoon, 2020 for a summary). Importantly, different offline tasks vary in their demands on working memory and in the degree to which they require integration of pragmatic information. However, research efforts aimed at addressing task effects and individual variation in this domain remain limited. Indeed, studies adopting online processing methods tend to report a local-reading advantage such that binding the SE to a local antecedent induces smaller processing costs compared to binding the SE to a LD antecedent (Dillon et al., 2016; Lyu & Kaiser, 2021).

Among bilingual speakers, to our knowledge, all existing studies have employed offline tasks. For example, C. Chen and Ionin (2023) used a picture-based truth-value judgment task to examine the acceptability of local and LD readings of PRO, SE, and SELF among two groups of Mandarin proficiency-matched L2 learners: L1-Korean and L1-English speakers. Their results showed that L1-Korean learners were more likely to accept local readings of PRO compared to both the L1-English learners and L1-dominant Mandarin speakers, suggesting cross-linguistic influence (CLI) from Korean, a language that permits locally bound PROs. In the comprehension of SELF, all three groups demonstrated a preference for local readings of *taziji*, although the L1-English group was somewhat more likely to reject its local readings. For SE, L1-dominant Mandarin speakers showed a numerical preference for local readings, while both L2 groups were significantly more likely to reject LD readings and accept local readings of *ziji*.

In a related study, C. Chen (2020) used the same methodology to compare the comprehension of all three pronoun forms among Mandarin-English heritage speakers (HSs) and L1-English L2 learners of Mandarin. The results showed that L2 learners were more likely to reject LD readings and accept local readings of PRO (in contrast to Chen & Ionin, 2023), SELF (partially consistent with Chen & Ionin, 2023), and SE (like Chen & Ionin, 2023). In contrast, HSs patterned more closely with L1-dominant speakers in their interpretation of PRO and SELF. However, they were less likely to accept the LD reading of *ziji* and more likely to accept its local reading, which was partially attributed to CLI from English.

The Present Study

The present study employs the web-based visual world eye-tracking paradigm to investigate the automatic, online processing of three types of Mandarin pronouns, *ta* (PRO), *ziji* (SE), and *taziji* (SELF), among Mandarin-English late adolescent HSs aged 14 to 18. Importantly, to avoid variable binding preferences of the SE across individuals and to bias a LD reading of the SE, we use only logophoric verbs in the experiment. This manipulation allows us to probe whether different pronoun types engage distinct resolution mechanisms during real-time processing. Specifically, we aim to address the following research questions (RQs):

RQ1: How do HSs process the three types of Mandarin pronouns? Do the pronouns elicit distinct processing patterns?

RQ2: Which (and how do) user-internal and user-external factors modulate IDs in HSs' processing? If so, do they do so differentially for different types of pronouns?

As described above, two user-internal factors are of particular interest in the present study: working memory (WM) and inhibition. As for user-external factor, we focus on HL input quantity. Our focus on late adolescence offers a particularly informative window for testing the effects of user-external and user-internal factors. Compared to early adolescence and childhood, late adolescence is marked by maturation of executive functions, enabling us to better distinguish between effects attributable to ongoing cognitive development and those driven by stable individual variability. Importantly, this age range is also when adolescents begin to exert greater autonomy over language use, making more independent choices about when, how, and with whom they engage in their HL. As a result, patterns of HL exposure and use (user-external factors) become more variable and personalized than in childhood, offering richer variation in external input that can be captured within an ID framework.

Predictions

Starting with RQ1, following C. Chen (2020)'s study with Mandarin-English adult HSs and the more broader HL processing literature (e.g., Fuchs, 2022; Hao, Chondrogianni, et al., 2024; Karaca et al., 2024), we predict that adolescent HSs will show distinct processing patterns across pronoun types. More specifically, they would show local binding preferences for *taziji* (SELF), and LD preferences for *ziji* (SE) and *ta* (PRO). Nevertheless, according to the Interface Hypothesis and the distance problem, HSs may show greater variability in their preferences during the processing of SE and PRO, but relatively robust local preferences for SELF. This is because SELF resolution is governed by narrow syntactic constraint (Binding Principle A) and involves local dependencies, whereas the resolution of SE and PRO requires the integration of

information across domains and the establishment of long-distance dependencies, making these structures more susceptible to input reduction and individual variation. The ID approach adopted in the present study allows us to move beyond a simple group-level analysis and ask why some HSs are more likely than others to exhibit local binding of SE and PRO despite the LD bias—driven by logophoric verbs in the case of SE and by Binding Principle B in the case of PRO—by examining the role of user-internal and user-external factors.

It is worth noting that local interpretations of the SE and PRO in the present study need not reflect the same underlying mechanisms. Because SE allows both local and LD binding in Mandarin and exhibits a default preference for local binding, local interpretations of SE in the present design may arise from multiple sources, including difficulties with interface/LD structures, reliance on a default local strategy, or cross-linguistic influence from English where reflexives are locally bound. In contrast, PRO does not permit local binding in either Mandarin or English; therefore, any local binding of PRO in the present study cannot be attributed to CLI or default strategies and is more likely to reflect difficulties associated with interface/LD integration.

For RQ2, we predict user-internal and user-external factors to play differential roles depending on the type of pronouns. With respect to user-internal factors, WM is expected to facilitate LD interpretations of SE and PRO, insofar as establishing LD dependencies requires maintaining and integrating multiple candidate antecedents and sources of information (e.g., Rule I, logophoric verb semantics) during online processing. In addition, inhibitory control is expected to modulate the extent to which SE is interpreted locally. Specifically, HSs with higher inhibitory ability may be better able to suppress English-like local binding strategies and/or Mandarin local binding preference, leading to higher likelihood of LD readings. Turning to user-external factors, despite it seemingly relatively intuitive that HL input will have a default, ubiquitous effect on HL processing, previously findings are mixed (Hao, Chondrogianni, et al., 2024; Hao, Kubota, et al., 2024; Karaca et al., 2024) and likely depend on other factors such as the specific domain of grammar and the time-course of its typical acquisition (Tsimplici, 2014). Two possibility emerges, i.e., HL input either modulates HL processing or not. Given the limited literature in this specific domain and age range, we treat the effect of HL input as exploratory, without strong directional predictions. However, given that the properties we examine, while inherently related, have independent time courses for full acquisition even in monolingual children (Chien & Lust, 2006) and in light of claims that reduced input would affect interface related structures more, it is possible, if not likely, that

input factors might have a greater influence over some properties (PRO and SE) examined herein than others (SELF).

Participants

In total, 125 eligible participants accessed the experimental platform and provided informed consent. Of these, 44 participants did not successfully complete the main visual-world eye-tracking task, primarily due to repeated calibration failures. An additional nine participants were excluded because their effective eye-tracking sampling rate during the main task was below 15Hz. Consequently, data from 72 Mandarin-English late adolescent HSs (14-18 y.o) who completed the study online were deemed of sufficient quality for analysis.

However, we further excluded 21 participants to control for variation in several user-level factors that have been shown to modulate HL development and use beyond the factors in focus in the current study. More specifically, we excluded three participants due to relatively low socioeconomic status, five participants who currently reside in Ireland, eight participants with low Mandarin proficiency measured by the Peabody Picture Vocabulary Test (PPTV) fourth edition (Dunn & Dunn, 2012), three participants who speak another language other than Mandarin and English (including other Chinese languages, e.g., Cantonese, Hokkien, etc.), and two participants who were exposed to the societal dominant language English after the age of three. The final sample included 51 participants (18 girls, mean age = 15.8 in years, SD = 1.7, min = 14 in years, max = 18 in years). Within the 51 participants, 16 currently reside in the UK, while the other 35 in the USA. All HSs were exposed to Mandarin from birth at home and to English before the age of 3 years. They were either born and raised (all second-generation immigrants; n = 17) or immigrated to their current residency before the age of 3 (first-generation immigrants; n = 34) with a mean age onset of acquisition of English 11.57 in months (SD = 7.96, min = 0, max = 27). All participants had exposure to formal instruction in Mandarin (e.g., via Saturday Schools, tuition, etc.).

Baseline Tasks

Language Background Questionnaire. To collect participants' language background and demographic information, we administered the Quantifying Bilingual Experience (Q-BEx) questionnaire (De Cat et al., 2023). Q-BEx is a validated, user-friendly online instrument designed to quantify multilingual language experience. We included all mandatory modules and optional modules, except for the detailed attitudinal module. As a result, we obtained a comprehensive assessment of HSs' language exposure and use, self-rated proficiency, richness

of linguistic experiences, and language mixing patterns. The questionnaire provides four composite scores that serve as proxies for HL use and exposure. Specifically, we derived two key measures: current HL exposure and use (Mean = 0.41, SD = 0.11, Range = 0.07-0.60) and cumulative HL exposure and use (Mean = 80.53, SD = 30.61, Range = 11.70-118.38), by summing the respective exposure and use components. These aggregated scores were used as continuous user-external variables in subsequent analyses.

English and Mandarin Receptive Vocabulary. We administered the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2012) to assess participants' receptive vocabulary ability in both English and Mandarin. Form A was administered in English, and Form B was translated from English into Mandarin and administered in Mandarin. Because the PPVT-4 is neither available nor normed for Mandarin and was not designed or normed for bilingual populations even in its English version, we report raw scores only, which were used solely for participant screening and exclusion purposes. For the same reasons, we caution readers against interpreting PPVT scores as direct measures of language proficiency or comparing English and Mandarin vocabulary scores within the same participant. Nevertheless, test administration followed the procedures outlined in the PPVT-4 manual, including age-appropriate starting items, ceiling rules, and termination criteria based on error counts. The final sample has a mean PPVT score of 139 in English (SD = 8.08, min = 115, max = 152) and 124 in Mandarin (SD = 7.44, min = 106, max = 151).

Flanker/No-Go Task. To examine inhibitory control, participants were tested on an engaging variant of a flanker task that also includes a Go and No-Go component (Woodard et al., 2016). In this task, a Go trial requires the participants to press a key on the keyboard ("Z" for left and "M" for right) in accordance with the direction of the middle fish that is surrounded by two flanker fishes on its left and two on its right. There are two conditions in the Go trials, the congruent condition (30 trials) and the incongruent condition (30 trials). In the congruent conditions, the flanker fishes face the same direction as the middle fish. In the incongruent conditions, the flanker fishes face the opposite direction to the middle fish. In a No-Go trial (30 trials), the middle fish is surrounded by fishbowls. The participants were instructed to refrain from responding. We counterbalanced the direction of the middle fish. Each trial was preceded by a 1000ms fixation cross in the middle of the screen and each trial lasted till a response was recorded or until a maximum display time of 5000ms was reached. Reaction time and accuracy were recorded. We calculated the No-Go cost (mean = -0.16, SD = 0.18, min = -0.33, max = 0.33) as a proxy for inhibitory control by taking the average accuracy difference between

omission errors on Go-trials and hits on No-Go trials. This No-Go cost score was carried forward into subsequent analyses, with higher values indexing weaker inhibitory control.

Working Memory Task. We used a spatial sequence working memory task inspired by the Alloway Working Memory Assessment (Alloway et al., 2008) to assess participants' nonverbal visuospatial working memory (WM). In this task, participants were instructed to help a forgetful alien return home by recalling, in reverse order, the sequence of squares the alien had walked through in a 4×4 matrix. In each trial, the alien randomly moved across one or more squares in the matrix. In the first block, the alien walked through a single square; in each subsequent block, the number of visited squares increased by one. Each block consisted of six trials, with a maximum of eight blocks in total. Scoring followed the standard procedure outlined in the Alloway assessment. If a participant responded correctly to the first four trials in a block, they automatically progressed to the next block and were awarded the full six points for that block. The task was terminated once the participant responded incorrectly to three trials within the same block. The mean score of the task reached 22.69 (SD = 4.76, min = 15, max = 30). This score was carried forward into the analyses as an index of working memory (WM) capacity, with higher scores indicating greater WM capacity.

The Visual World Eye-Tracking Experiment. The visual world eye-tracking paradigm was adopted to examine participants' online processing. In the task, participants listened to sentences while viewing three pictures (of potential referents) on the screen (Figure 1). We embedded the pronouns in genitive forms by adding the genitive marker *de* after the pronouns and we included three experimental conditions (i.e., PRO, SE vs. SELF). Additionally, we included a condition with full noun phrases (NPs) followed by *de* marker as a control condition to make sure participants understand genitives. Each condition had nine trials, giving rise to 27 experimental trials and 9 control trials. All experimental sentences followed the same format: This morning/last night + Long Distance referent NP + Main Verb + Local referent NP + Embedded Verb + PRO/SE/SELF/control NP + *de* + NP. We chose three logophoric verbs (*xiangrang* “want somebody do something”, *yaoqiu* “demand somebody do something”, and *mengjian* “dream about somebody doing something”) as the main verbs and each appeared three times per condition. For the embedded verbs and the *de* NP pairs, we chose three pairs, i.e., *liang ... de tiwen* “checking someone’s temperature”, *mo ... de erdu* “touch someone’s ear”, and *la ... de weiba* “pull someone’s tail”, and each appeared three times per condition. For all referent NPs (including for the control condition NPs), frequently used disyllabic animals were adopted, and each animal appeared the same time either as the LD referent, the local referent or the third referent (either not mentioned or the control condition NP). The

positions of the NPs were counterbalanced such that the LD referent, local referent and a third potential referent appeared equal numbers of time on Top, Bottom Left and Bottom Right positions. To avoid item-specific effects, we created four lists such that each item appeared once as PRO, SE, SELF or NP condition. For example, figure 1 as a visual scene was accompanied by sentence (5a) in List A, (5b) in List B, (5c) in List C, and (5d) in List D.

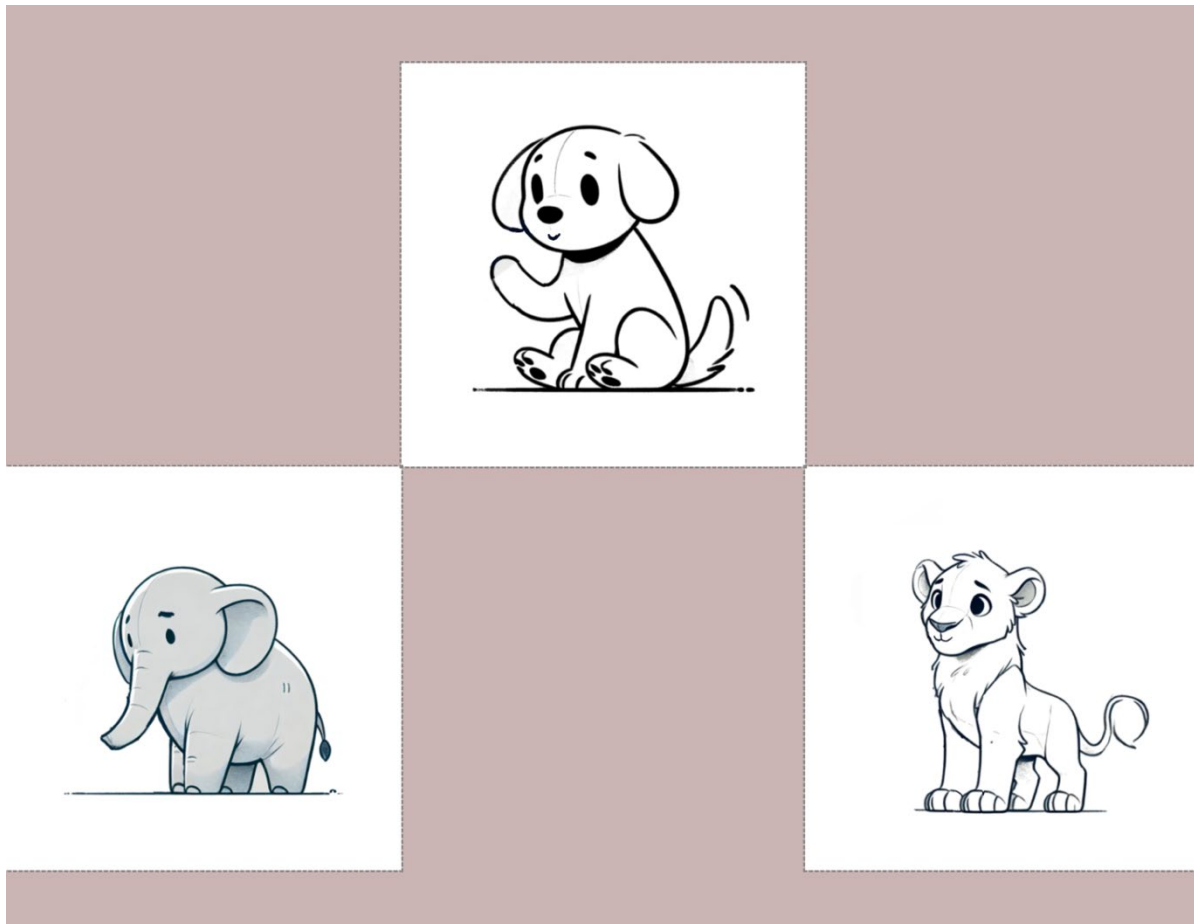


Figure 1. Example of a visual scene in the visual world eye-tracking experiment.

- (5) a jintianzaoshang daxiang yaoqiu xiaogou liang ta de tiwen
 This morning elephant demand dog check PRO de temperature
 “This morning, the elephant demanded the dog to check his temperature.”
- b jintianzaoshang daxiang yaoqiu xiaogou liang ziji de tiwen
 This morning elephant demand dog check SE de temperature
 “This morning, the elephant demanded the dog to check his temperature/the temperature of himself.”
- c jintianzaoshang daxiang yaoqiu xiaogou liang taziji de tiwen
 This morning elephant demand dog check SELF de temperature

“This morning, the elephant demanded the dog to check the temperature of himself.”

d jintianzaoshang daxiang yaoqiu xiaogou liang shizi de tiwen
This morning elephant demand dog check lion de temperature

“This morning, the elephant demanded the dog to check the lion’s temperature.”

The experiment also included 18 filler trials where the processing of relative clauses was the focus. Relative clauses were chosen as they allow us to make sure two animate referents can be mentioned while another one can be inferred, following Y. T. Huang et al. (2013)’s design. As such, the visual scenes for the relative clauses are comparable to the ones used for pronoun processing in terms of animacy and number of potential referents. For example, for the filler sentence “The fish that catches the lion is singing.”, the visual scene consisted of a fish, a lion and a potential patient of the action not mentioned in the sentence, e.g., a shrimp, or a potential agent of the action not mentioned, e.g., a crab.

To ensure that the participants paid attention to the sentences and to make sure the participants understood genitives, the control trials along with another random 15 trials included an offline comprehension check in the format of picture sentence verification (no participant was removed due to low/below-chance offline comprehension accuracy). In the comprehension check, participants were shown a picture that either matched the sentence or not and asked to press the “Z” key on the keyboard if they matched or the “M” key if not. This comprehension check was embedded in another alien game. We instructed the participants to listen to the sentences and look at some pictures. We also informed them that occasionally, an alien will try to draw the event described in the sentence. They were instructed that when this happens, their task was to decide whether the alien’s drawing matched the sentence or not by pressing “Z” or “M”. All trials appeared in a completely random order for each participant. Each trial began with a 1,500-ms display of the visual scene, followed by the auditory experimental sentence.

Auditory stimuli were recorded by a male Mandarin L1-dominant user in a sound-proof booth. Experimental stimuli were constructed by concatenating extracted tokens of This morning/last night + Long Distance referent NP + Main Verb + Local referent NP + Embedded Verb + PRO/SE/SELF/control NP + *de* + NP. All recordings were produced with neutral prosody, and no systematic prosodic manipulation (e.g., stress) were implemented on any segments. Two L1-dominant Mandarin users checked the naturalness of all stimuli. The duration of all parts but the PRO/SE/SELF/control NP were held constant across all items. Importantly, the duration of *de* + NP was exactly 1,200ms in each experimental item. This 1,200ms period constitutes the critical region for analysis.

Procedure

All participants took part in the study at their homes. We implemented all tasks with Gorilla on a webpage (Anwyl-Irvine et al., 2020), which utilizes WebGazer.js (Papoutsaki et al., 2016) to run webcam-based eye-tracking. To minimize any carry over effect between the Mandarin and English vocabulary test, all participants completed all tasks in the follow sequence: the Eye-tracking task, Mandarin PPVT, Flanker/No Go Task, Working Memory Task, English PPVT and the Q-Bex. The whole experiment lasts around approx. 65 mins. The study was approved by the institutional ethics committee. All participants were informed of their ethical rights of participation in written form, prior to the experiment. Before any tasks, participants were asked to check boxes on the webpage to give consent for their participation.

Prior to participation, participants received an introduction video accompanied by written instructions in both Mandarin and English, detailing how participants could help in optimizing data quality (e.g., close all other applications and webpages except for the experiment page; maximize ambient lighting, etc.). Participants were additionally provided with both video and written instructions on how to complete the calibration procedure. Eye-tracking calibration employed a 9-point calibration routine. Recalibration was performed every nine trials or every five minutes, whichever occurred first. For each calibration phase, participants were allowed up to three attempts. An attempt was classified as unsuccessful if at least two out of the nine calibration points failed to calibrate successfully. Furthermore, to minimize system lag and reduce computational load, eye-tracking data were recorded only from the onset of the LD NPs. As a result, eye-gaze data during the 1,500-ms preview window and during the initial temporal adverbial segment (e.g., “last night” / “this morning”) were not recorded. These quality optimization procedures gave us a mean effective sampling rate of 30.6Hz (SD = 9.3, Range = 15-60).

Results

For plotting and data analyses, we resampled the eye-movement data into 50ms time bins. Given the mean sampling rate, this bin width minimizes empty bins while avoiding excessive aggregation of multiple samples within a single bin. This resampling yielded 24 time bins (data points) within the critical time window. To ensure data quality, we excluded trials with more than 50% of invalid data points (e.g., out of bounds), resulting in 1,285 trials retained out of a total of 1,479 trials. Figure 2 illustrates the difference in mean fixation proportion to the long distance (LD) referent minus the mean fixation proportion to the local referent throughout the

course of a trial. As such, a positive value indicates more looks to the LD referent over the local referent and a negative value indicates more looks to the local referent over the LD referent. The dotted vertical line indicates the onset of the critical time window (the onset of the genitive marker *de*). As we centered the time information with reference to the onset of the genitive marker and as different pronouns have different lengths, the onset of each condition differs in the figure. Visual inspection suggests that at the group level (RQ 1), HSs preferred LD referent over local referent after hearing the PRO (LD-advantage score: Mean = 10.88, SD = 4.37, Range = 1-21) and the SE (LD-advantage score: Mean = 6.76, SD = 6.65, Range = -13-21). However, a preference for the local over LD referent was observed after HSs hearing the SELF (LD-advantage score: Mean = -7.40, SD = 3.89, Range = -24-0).

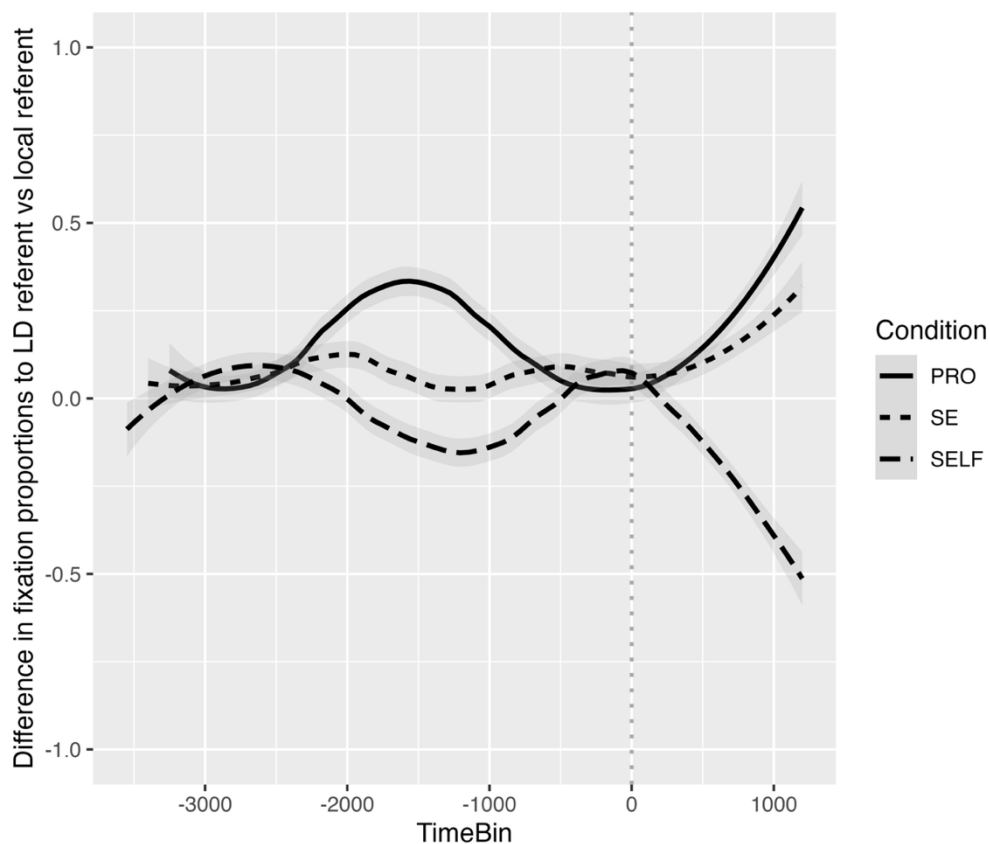


Figure 2. Difference in proportion fixations to LD vs. local referent by Condition.

To statistically account for the results, we calculated the long distance (LD) advantage score (over local referent) within the critical time window. This LD-advantage score was calculated by subtracting the number of 50ms time bins within the critical time window that contained looks to the local referent from the number of 50ms time bins that contained looks to the LD referent. We did not adjust for the 200ms needed to initiate ballistic eye movement in response to an acoustic stimulus because we do not make a distinction between prediction and integration. In contrast, we are interested in how pronouns are interpreted online in general.

For statistical analyses, general linear mixed effect regressions were carried out with the lme4 package (Bates et al., 2015) in R (R Core Team, 2018). Our statistical modelling, whenever possible, follows a confirmatory approach that is subjective and theory-driven (McElreath, 2020), where only theory-driven fixed effect factors were included. For random effects, we included the maximal random effects justified by the design where possible (Barr et al., 2013), i.e., by-subject and by-items random intercepts, as well as by-subject and by-item random slopes for Condition. When the maximal model failed to converge, we iteratively simplified random effect structures until convergence was achieved, i.e., removing random effect(s) accounting for the least variance.

For RQ1 (Do the pronouns elicit distinct processing patterns?), the maximal converged model was derived via the R syntax: `lmer(LD_Adv ~ Condition + (1+Condition|Participant) + (1|Item)`, with the fixed effect Condition treatment coded and SE as the referenced level (SE vs. PRO and SELF). The model suggests that (1) the SE condition induced significantly more looks to the LD referent over the local referent (the intercept is significantly larger than zero: Estimate = 7.41, SE = 0.51, CI [6.40, 8.42], $t = 14.42$, $p < 0.001$); (2) the PRO condition induced more looks to the LD referent compared to the SE condition (Estimate = 3.46, SE = 0.48, CI [2.51, 4.41], $t = 7.15$, $p < 0.001$); and (3) the SELF condition induced more looks to the local referent compared to the SE condition (Estimate = -15.38, SE = 0.74, CI [-16.83, -13.92], $t = -20.76$, $p < 0.001$). As post hoc analyses, we reran the model with referenced reference level for both PRO and SELF to examine HSs' preference for LD and local referent for each condition. The results suggest that the PRO condition induced significantly more looks to the LD referent over the local referent (the intercept is significantly larger than zero: Estimate = 10.87, SE = 0.35, CI [10.18, 11.56], $t = 31.00$, $p < 0.001$). In contrast, the SELF condition induced significantly more looks to the local referent over the LD referent (the intercept is significantly smaller than zero: Estimate = -7.96, SE = 0.48, CI [-8.91, -7.01], $t = -16.42$, $p < 0.001$).

For RQ2 (RQ2: Which (and how do) user-internal and user-external factors modulate IDs in HSs' processing?), we firstly examined the correlations among all individual-internal and individual-external factors to avoid multicollinearity in statistical modeling. This step identified strong correlations between Mandarin PPVT score and all other variables, between current HL exposure and use and cumulative HL exposure and use, among others (see the R script on OSF for more information). As such, aiming to statistically account for the effects of theoretical interests, we decided to include only current HL exposure and use (centered), working memory (WM; centered) and inhibition (No Go Cost; centered) as fixed effects interacting with Condition (treatment coded). Given the quantity of data at our disposal,

we did not include interaction terms in the models among these background factors. The final converged maximal model has the R syntax of `lmer(LD_Adv ~ Condition*(Current_Exposure_Use_Mandarin_c + WM_c + NoGo_Cost_c) + (1|Participant) + (1|Item))`. We calculated the Variance Inflation Factor (VIF) to ensure that the maximal model does not violate multicollinearity principle (all VIFs < 2). Table 1 summarizes the statistical output where SE is treated as the reference level for the Condition variable (SE vs. PRO, SELF). It is important to note that as the categorical variable was treatment-coded, the effects reported in the statistical table represent simple effects, i.e., the effect of a variable at a given level relative to the reference level, rather than main effects.

Table 1. The model with Condition (SE as the reference level) interacting with Current HL Exposure and Use, WM, and NoGo Cost as fixed effects.

<i>Predictors</i>	<i>Estimates</i>	<i>SE</i>	<i>CI</i>	<i>t</i>	<i>p</i>
(Intercept)	7.41	0.29	6.83 – 7.98	25.36	<0.001
SELF	-15.42	0.41	-16.23 – -14.61	-37.39	<0.001
PRO	3.46	0.36	2.76 – 4.17	9.62	<0.001
Current HL Exposure Use	0.50	0.26	-0.00 – 1.00	1.95	0.052
WM	1.78	0.25	1.30 – 2.27	7.25	<0.001
NoGo Cost	-2.13	0.25	-2.63 – -1.63	-8.41	<0.001
SELF × Current HL Exposure Use	-2.62	0.34	-3.30 – -1.95	-7.62	<0.001
PRO × Current HL Exposure Use	-0.66	0.30	-1.24 – -0.08	-2.24	0.025
SELF × WM	-2.18	0.35	-2.86 – -1.50	-6.26	<0.001
PRO × WM	0.14	0.29	-0.42 – 0.70	0.48	0.628
SELF × NoGo Cost	1.99	0.35	1.31 – 2.67	5.74	<0.001
PRO × NoGo Cost	2.13	0.29	1.55 – 2.70	7.27	<0.001

This model suggests that individual HSs' look patterns for the SE condition were modulated by WM and NoGo Cost but not by Current HL Exposure and Use. More specifically, with the increase of WM, there is an increase in looks to the LD referent and with the increase of NoGo Cost, there is a decrease in looks to the LD referent. To unpack the significant interaction terms between Condition and WM, NoGo Cost and Current HL Exposure and Use, we ran post hoc analyses by conducting models with all possible combinations of reference levels for the categorical variable Condition. As can also be seen in figure 3, post hoc

analyses reveal that participants' look pattern for the PRO condition was modulated by WM where participants with higher WM load were more likely to look at the LD referent (Estimate = 1.92, SE = 0.24, CI [1.45, 2.39], $t = 8.07$, $p < 0.001$). However, look pattern for the PRO condition was not modulated by Current HL Exposure and Use (Estimate = -0.17, SE = 0.25, CI [-0.65, 0.32], $t = -0.67$, $p = 0.51$) nor by NoGo Cost (Estimate = -0.01, SE = 0.24, CI [-0.47, 0.47], $t = -0.02$, $p = 0.98$). Look pattern for the SELF condition was not modulated by WM (Estimate = -0.40, SE = 0.31, CI [-1.01, 0.21], $t = -1.27$, $p = 0.20$) nor by NoGo Cost (Estimate = -0.14, SE = 0.30, CI [-0.73, 0.46], $t = -0.45$, $p = 0.65$) but was modulated by Current HL Exposure and Use such that participants with more HL exposure and use were more likely to look at the local referent (Estimate = -2.13, SE = 0.30, CI [-2.72, -1.53], $t = -6.99$, $p < 0.001$).

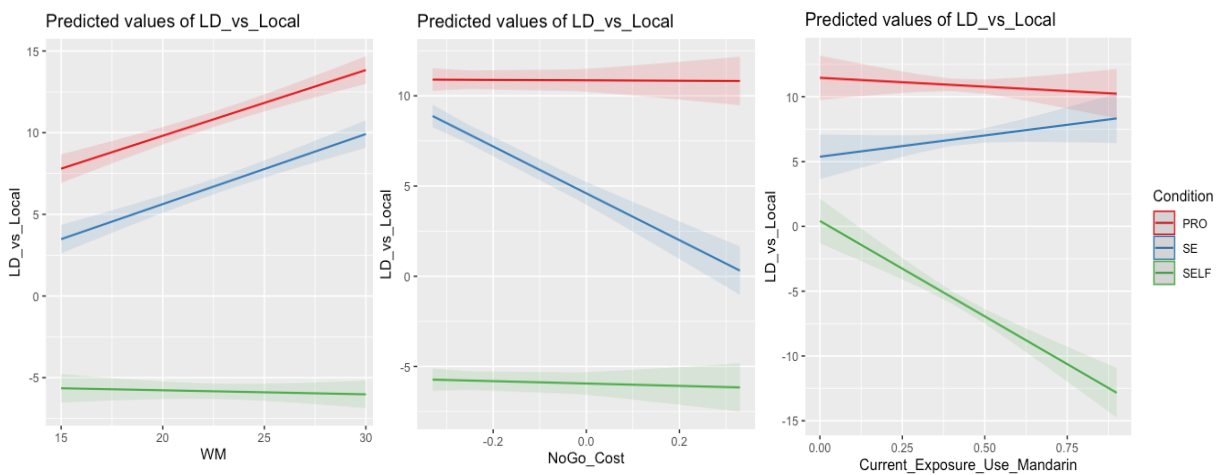


Figure 3. Effect of WM (left), Inhibition (mid) and Current Exposure and Use of Mandarin on LD advantage

Discussion

The present study investigated how Mandarin-English HSs interpret different types of pronouns in real time and how linguistic-level and individual-level factors shape this process. We focused on a group of late adolescent HSs, an age group that are underrepresented in the literature. Using a web-based visual world eye-tracking paradigm, we examined the online interpretation of three Mandarin pronouns, i.e., *ta* (PRO), *ziji* (SE), and *taziji* (SELF). Importantly, while SELF is governed by narrow syntax, interpreting PRO and SE requires integrating different information sources, with PRO requiring syntax-discourse integration and SE requiring syntax-semantics-discourse integration.

Starting with group-level performance across pronoun types (RQ1), we observed clear distinctions in pronoun processing patterns. Specifically, HSs interpreted SELF as referring to the local antecedent, whereas PRO and SE were interpreted as referring to the long-distance

(LD) antecedent. These findings align with those reported by C. Chen (2020), who found L1-dominant-like offline interpretations of PRO and SELF among adult HSs. Our results combine with that C. Chen (2020) to suggest that Mandarin HSs successfully process and interpret pronouns, even when pronominal resolution requires establishing complex linguistic dependencies, such as those involving long-distance antecedents and are at the interface between syntax (semantics) and discourse (as in PRO and SE). This observation is particularly noteworthy in light of claims that linguistic dependencies, especially long-distance ones, constitute a vulnerable domain in HL grammars (Polinsky & Scontras, 2020). Moreover, under the IH (Sorace, 2011), structures that require integration across grammatical interfaces between syntax and discourse are predicted to be especially prone to attrition or arrested development in bilingual populations. Both PRO and SE fall into this category, as their resolution depends on information at the syntax-(semantics-)discourse interface.

The absence of evidence for such vulnerability in our data, therefore, challenges the scope of these theoretical predictions, highlighting the need to better understand the conditions under which interface-dependent phenomena/long-distance dependencies may or may not be vulnerable in HSs (Daskalaki et al., 2019; Leal et al., 2014). One plausible explanation for this discrepancy is methodological. Whereas previous studies reporting vulnerability in pronominal interpretation among HSs have primarily used offline comprehension or production tasks and focused on adult HSs (e.g., Kim & Yoon, 2020), our study employed a real-time eye-tracking paradigm and targeted late-adolescent HSs. The findings thus underscore the need for future research that adopts a developmental perspective and systematically varies task modality to capture a more comprehensive picture of HL bilingual processing and representation across the lifespan (Fuchs, 2022).

Additionally, the current study found that PRO elicited more looks to the LD referent than SE, suggesting stronger and more consistent LD resolution for PRO across participants. Strikingly, all LD advantage scores for the PRO condition were positive, indicating that every participant consistently interpreted PRO as referring to the LD antecedent across all items. In contrast, the LD advantage scores for SE showed much more variability. This suggests that although there was a group-level preference for LD interpretations of SE, individual participants occasionally interpreted SE as referring to the local antecedent. This variability is not entirely unexpected, even though all the verbs used in the study were logophoric, biasing LD-readings of the SE. This may reflect differences in the linguistic architecture supporting each pronoun. PRO resolution relies on syntax-discourse interface mechanisms. In contrast, SE resolution requires integration across multiple interfaces (syntax, semantics, and discourse),

potentially making it more susceptible to individual variation whereby HSs may weigh sources of information differently depending on their individual language experiences and cognitive resources (Hao, Kubota, et al., 2024; Hao et al., 2025; Kim & Yoon, 2020). These findings underscore the importance of examining not only group-level trends but also participant-level variability in HL bilingualism (De Houwer, 2023; Paradis, 2023; Rothman et al., 2023), a point we now turn to in the discussion of RQ2.

Our second research question (RQ2) investigated which user-internal (i.e., WM, inhibition) and user-external (i.e., current HL exposure and use) factors modulate IDs in HSs' pronoun processing and whether these effects differ across pronoun types. The results reveal distinct patterns of modulation for each pronoun. For SELF, only the user-external factor of current HL exposure and use significantly modulated participants' look patterns: HSs with greater current exposure to and use of the HL were more likely to fixate on the local antecedent. In contrast, for PRO, only the user-internal factor of WM, but not inhibition or HL exposure and use, predicted processing behavior. That is, participants with higher WM capacity showed stronger preferences for the LD antecedent. Lastly, for SE, both WM and inhibition modulated look patterns: participants with higher WM capacity and stronger inhibitory control (indicated by lower NoGo cost) showed a greater tendency to fixate on the LD antecedent.

The role of user-internal factors aligns well with our predictions and broader cognitive accounts of sentence processing. Specifically, WM emerged as a key modulator of successful LD resolution for both PRO and SE, but not for SELF. This pattern suggests that pronouns that require resolution across longer syntactic distances or across multiple domains (e.g., syntax, discourse, semantics) impose greater cognitive demands and thus rely more heavily on WM resources. In contrast, processing of SELF, which is governed primarily by narrow syntactic constraints and typically resolved locally, does not appear to require substantial WM resources. This accords well with findings from Bice and Kroll (2021), who also reported no WM effects on morphosyntactic processing (subject-verb agreement) among HSs, reinforcing the view that local dependencies/narrow syntax may not engage domain-general cognitive systems to the same extent compared to long distance dependencies/interface structures.

Unlike PRO, SE exhibited sensitivity to both WM and inhibition, suggesting that processing SE may place demands not only on memory resources but also on participants' ability to manage competing interpretations. One possibility is that inhibition is involved in suppressing CLI. English lacks a direct equivalent of Mandarin SE, which might lead HSs to map English reflexives onto Mandarin SE, giving rise to a preference for local interpretations of the SE. However, this does not mean that Mandarin HSs have a reduced inventory of

pronouns (PRO vs. reflexives), a possibility suggested by Polinsky and Scontras (2021), as the current study does show a clear distinction in how HSs process the three different pronouns. Alternatively, or perhaps additionally, inhibition may be required to suppress a dominant or default preference for local binding of the SE, even when the semantic cues from the logophoric verb bias a LD reading. This is supported by prior research suggesting that local interpretations of the SE are more accessible and less costly for processing (Dillon et al., 2016; Lyu & Kaiser, 2021). Under such an account, participants with better inhibitory control were more successful in suppressing the local (default) interpretation especially when it conflicted with the logophoric semantics of the main verbs.

The fact that inhibition modulated SE but not PRO, despite both requiring LD interpretation, further suggests that SE may involve more interpretive conflict. An open question, however, concerns which of these mechanisms, i.e., CLI vs. default local binding preference, is more influential in shaping SE interpretation among HSs. Future studies could address this question by directly manipulating verb type (e.g., logophoric vs. generic), thereby testing whether the availability of strong semantic cues reduces inhibitory demands in SE interpretation. Additionally, examining the processing of other long-distance dependencies that differ in across-language similarities/differences and within-language defaults may help dissociate CLI effects from interpretive biases.

Lastly, regarding the effect of the user-external factor, i.e., current HL exposure and use, we found that it modulated look patterns only when HSs processed the SELF, but not the SE or the PRO. While it is seemingly intuitive that more HL exposure and use would lead to better performance in general as observed in offline tasks (Chondrogianni & Daskalaki, 2023; Daskalaki et al., 2019; Kubota et al., 2025; Paradis et al., 2020; Soto-Corominas et al., 2021), findings from online processing studies have been mixed. HL input effects have been reported in adult HSs (e.g., Hao, Kubota, et al., 2024; Hao et al., 2025), yet such effects have been shown to be absent in children (e.g., Hao, Chondrogianni et al., 2024). While one interpretation is that HL input effects are developmentally mediated (i.e., the older one gets, possibly according with more distance in multiple senses the average HS has to the HL, IDs in exposure become (more) deterministic), the current findings suggest a more nuanced possibility: HL input may have selective effects on real-time processing depending on the grammatical domain.

Specifically, we propose that structures involving interface-level integration and/or long-distance dependencies (the SE and PRO), place greater demands on domain-general cognitive resources. As such, their processing may rely less on the amount of HL input alone, and more on individual cognitive capacities like WM and inhibition. In contrast, narrow

syntactic structures, such as SELF, which is constrained by Binding Principle A and requires local binding, may benefit more directly and robustly from increased HL exposure and use, as these structures are more rule-governed, frequent in input, and less cognitively taxing.

This interpretation, however, stands in contrast to the predictions of the IH (Sorace, 2011), which posits that interface phenomena are more vulnerable to variability and should be more sensitive to input (reduction). One possible resolution is to distinguish between representation and processing efficiency, especially when most studies supporting the IH come from offline measures. It may be that increased HL use enhances the automaticity and efficiency of processing well-established syntactic representations (e.g., Hao, Chondrogianni, et al., 2024; Hao, Kubota, et al., 2024) but is less effective at resolving the more variable and inferential demands of interface phenomena, where cognitive effort, rather than input frequency, may be the bottleneck.

There is evidence from the present study that supports this view. When we examined individual participants' LD vs. local antecedent preferences for SELF across items, we found that all participants consistently interpreted SELF as referring to the local antecedent as LD advantage scores were uniformly negative. This suggests categorical application of Binding Principle A. The role of HL exposure and use here may not have influenced which interpretation participants arrived at, but rather how efficiently they processed and resolved the dependency in real time. In other words, HL exposure may have facilitated more rapid retrieval or application of binding constraints, even when interpretive outcomes were uniform across participants. Future research could test this by manipulating HL use, comparing real-time processing and interpretive accuracy across pronoun types, using longitudinal or training designs, and varying verb semantics to isolate input effects from cognitive demands.

Overall, the current findings strongly suggest that not all interface phenomena/long-distance dependencies are equally vulnerable, nor do they uniformly respond to input variation or cognitive factors (Leal et al., 2014). Instead, this study highlights the nuanced role cognitive and experiential factors play in shaping real-time pronoun processing among adolescent HSs.

Data Availability Statement: Supplementary materials, including the full experimental lists, and the data that support the findings of this study are openly available in OSF at https://osf.io/xs23z/overview?view_only=5d3bcd517f8547cc98e53115d449dc0a

References:

- Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology, 28*(7), 725–734. <https://doi.org/10.1080/01443410802243828>
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bayram, F., Pisa, G. D., Rothman, J., & Slabakova, R. (2021). Current Trends and Emerging Methodologies in Charting Heritage Language Grammars. In S. Montrul & M. Polinsky (Eds.), *The Cambridge Handbook of Heritage Languages and Linguistics* (1st ed., pp. 545–578). Cambridge University Press. <https://doi.org/10.1017/9781108766340.025>
- Bice, K., & Kroll, J. F. (2021). Grammatical processing in two languages: How individual differences in language experience and cognitive abilities shape comprehension in heritage bilinguals. *Journal of Neurolinguistics, 58*, 100963. <https://doi.org/10.1016/j.jneuroling.2020.100963>
- Chen, C. (2020). The Acquisition of Mandarin Relative Clauses and Binding by Heritage Speakers and Second Language Learners. *Proceedings of the 44th Boston University Conference on Language Development, 77–90*.

- Chen, C., & Ionin, T. (2023). Interpretation of Mandarin pronouns and reflexives by L1-Korean and L1-English learners of Mandarin. *Second Language Research*, 39(4), 941–968. <https://doi.org/10.1177/02676583221103744>
- Chien, Y.-C. (1992). Theoretical Implications of the Principles and Parameters Model for Language Acquisition in Chinese. In *Advances in Psychology* (Vol. 90, pp. 313–345). Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)61896-8](https://doi.org/10.1016/S0166-4115(08)61896-8)
- Chien, Y.-C., & Lust, B. (2006). Chinese children’s knowledge of the Binding Principles. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *The Handbook of East Asian Psycholinguistics* (1st ed., pp. 23–38). Cambridge University Press. <https://doi.org/10.1017/CBO9780511550751.004>
- Chien, Y.-C., & Wexler, K. (1990). Children’s knowledge of locality conditions in Binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition*, 1(3), 225–295. https://doi.org/10.1207/s15327817la0103_2
- Chomsky, N. (1993). *Lectures on government and binding: The Pisa lectures* (7th ed). Mouton de Gruyter.
- Chondrogianni, V. (2023). Chapter 12. Cross-linguistic influences in bilingual morphosyntactic acquisition. In I. Elgort, A. Siyanova-Chanturia, & M. Brysbaert (Eds.), *Bilingual Processing and Acquisition* (Vol. 16, pp. 294–315). John Benjamins Publishing Company. <https://doi.org/10.1075/bpa.16.12cho>
- Chondrogianni, V., & Daskalaki, E. (2023). Heritage language use in the country of residence matters for language maintenance, but short visits to the homeland can boost heritage language outcomes. *Frontiers in Language Sciences*, 2, 1230408. <https://doi.org/10.3389/flang.2023.1230408>

- Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language*, 65(2), 128–144. <https://doi.org/10.1016/j.jml.2011.04.007>
- Cole, P., Hermon, G., & Huang, C.-T. J. (2006). Long-Distance Binding in Asian Languages. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (pp. 21–84). Blackwell Publishing. <https://doi.org/10.1002/9780470996591.ch39>
- Cummings, I. (2017). Parsing and Working Memory in Bilingual Sentence Processing. *Bilingualism: Language and Cognition*, 20(4), 659–678. <https://doi.org/10.1017/S1366728916000675>
- Daskalaki, E., Chondrogianni, V., Blom, E., Argyri, F., & Paradis, J. (2019). Input effects across domains: The case of Greek subjects in child heritage language. *Second Language Research*, 35(3), 421–445. <https://doi.org/10.1177/0267658318787231>
- De Cat, C., Kaščelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., & The Q-BEx Consortium. (2023). How to quantify bilingual experience? Findings from a Delphi consensus survey. *Bilingualism: Language and Cognition*, 26(1), 112–124. <https://doi.org/10.1017/S1366728922000359>
- De Houwer, A. (2023). The danger of bilingual–monolingual comparisons in applied psycholinguistic research. *Applied Psycholinguistics*, 44(3), 343–357. <https://doi.org/10.1017/S014271642200042X>
- Dillon, B., Chow, W.-Y., & Xiang, M. (2016). The Relationship Between Anaphor Features and Antecedent Retrieval: Comparing Mandarin Ziji and Ta-Ziji. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01966>
- Dunn, L. M., & Dunn, D. M. (2012). *Peabody Picture Vocabulary Test—Fourth Edition* [Dataset]. <https://doi.org/10.1037/t15144-000>

- Fuchs, Z. (2022). Eyetracking evidence for heritage speakers' access to abstract syntactic agreement features in real-time processing. *Frontiers in Psychology, 13*, 960376. <https://doi.org/10.3389/fpsyg.2022.960376>
- Grodzinsky, Y., & Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry, 24*, 69–102.
- Hao, J., & Chondrogianni, V. (2023). Comprehension and production of non-canonical word orders in Mandarin-speaking child heritage speakers. *Linguistic Approaches to Bilingualism, 13*(4), 468–499. <https://doi.org/10.1075/lab.20096.hao>
- Hao, J., Chondrogianni, V., & Sturt, P. (2024). Heritage language development and processing: Non-canonical word orders in Mandarin–English child heritage speakers. *Bilingualism: Language and Cognition, 27*(3), 334–349. <https://doi.org/10.1017/S1366728923000639>
- Hao, J., Kubota, M., Bayram, F., González Alonso, J., Grüter, T., Li, M., & Rothman, J. (2024). Schooling and home language usage matter in heritage bilingual processing: Sortal classifiers in Mandarin. *Second Language Research, Online First*. <https://doi.org/10.1177/02676583241270900>
- Hao, J., Rossi, E., Nakamura, M., Luque, A., & Rothman, J. (2025). Individual differences matter in heritage language bilingual processing: An electroencephalography (EEG) study of grammatical gender. *Studies in Second Language Acquisition, 1–20*. <https://doi.org/10.1017/S0272263125101149>
- Huang, C. T. J., & Liu, C. L. (2000). Logophoricity, attitudes, and Ziji at the interface. *Syntax and Semantics, 33*, 141–195.
- Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language, 69*(4), 589–606. <https://doi.org/10.1016/j.jml.2013.08.002>

- Karaca, F., Brouwer, S., Unsworth, S., & Huettig, F. (2024). Morphosyntactic predictive processing in adult heritage speakers: Effects of cue availability and spoken and written language experience. *Language, Cognition and Neuroscience*, 39(1), 118–135. <https://doi.org/10.1080/23273798.2023.2254424>
- Kim, E. H., & Yoon, J. (2020). Experimental evidence supporting the overlapping distribution of core and exempt anaphors: Re-examination of long-distance bound *caki-casin* in Korean. *Linguistics*, 58(6), 1775–1806. <https://doi.org/10.1515/ling-2020-0233>
- Korkus, M.-L., & Vihman, V.-A. (2024). Adolescent Heritage Speakers: Morphosyntactic Divergence in Estonian Youth Language Usage in Sweden. *Languages*, 9(12), 366. <https://doi.org/10.3390/languages9120366>
- Kubota, M., Goto, Y., Kurokawa, S., Matsuoka, Y., Otani, M., & Rothman, J. (2025). Different variables hold varying significance from childhood to adolescence: Exploring individual differences in grammar development of Japanese heritage speakers. *Studies in Second Language Acquisition*, 47(1), 104–135. <https://doi.org/10.1017/S0272263124000615>
- Kupisch, T., & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, 22(5), 564–582. <https://doi.org/10.1177/1367006916654355>
- Leal, T., Rothman, J., & Slabakova, R. (2014). A Rare Structure at the Syntax-Discourse Interface: Heritage and Spanish-Dominant Native Speakers Weigh In. *Language Acquisition*, 21(4), 411–429. <https://doi.org/10.1080/10489223.2014.892946>
- Lyu, J., & Kaiser, E. (2021). Unpacking the blocking effect: Syntactic prominence and perspective-taking in antecedent retrieval in Mandarin Chinese. *Glossa: A Journal of General Linguistics*, 6(1). <https://doi.org/10.16995/glossa.5781>

- MacWhinney, B. (2018). A unified model of first and second language learning. In M. Hickmann, E. Veneziano, & H. Jisa (Eds.), *Trends in Language Acquisition Research* (Vol. 22, pp. 287–312). John Benjamins Publishing Company.
<https://doi.org/10.1075/tilar.22.15mac>
- McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN* (2nd edition). Chapman and Hall/CRC.
- Minkov, M., Kagan, O., Protassova, E., & Schwartz, M. (2019). Towards a Better Understanding of a Continuum of Heritage Language Proficiency: The Case of Adolescent Russian Heritage Speakers. *Heritage Language Journal*, 16(2), 211–237.
<https://doi.org/10.46538/hlj.16.2.5>
- Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge, United Kingdom : Cambridge University Press.
- Montrul, S., & Polinsky, M. (Eds.). (2021). Research Approaches to Heritage Languages: In *The Cambridge Handbook of Heritage Languages and Linguistics* (1st ed., pp. 373–578). Cambridge University Press. <https://doi.org/10.1017/9781108766340.017>
- Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). Webgazer: Scalable webcam eye tracking using user interactions. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 3839–3845.
- Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, 1–25.
<https://doi.org/10.1017/S0305000922000708>
- Paradis, J., Soto-Corominas, A., Chen, X., & Gottardo, A. (2020). How language environment, age, and cognitive capacity support the bilingual development of Syrian refugee children recently arrived in Canada. *Applied Psycholinguistics*, 41(6), 1255–1281. <https://doi.org/10.1017/S014271642000017X>

- Polinsky, M. (2018). Bilingual children and adult heritage speakers: The range of comparison. *International Journal of Bilingualism*, 22(5), 547–563.
<https://doi.org/10.1177/1367006916656048>
- Polinsky, M., & Scontras, G. (2020). Understanding heritage languages. *Bilingualism: Language and Cognition*, 23(1), 4–20. <https://doi.org/10.1017/S1366728919000245>
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reuland, E. (2006). Binding Theory. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (pp. 260–283). Blackwell Publishing.
<https://doi.org/10.1002/9780470996591.ch9>
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, 13(2), 155–163. <https://doi.org/10.1177/1367006909339814>
- Rothman, J., Bayram, F., DeLuca, V., Di Pisa, G., Duñabeitia, J. A., Gharibi, K., Hao, J., Kolb, N., Kubota, M., Kupisch, T., Laméris, T., Luque, A., Van Osch, B., Pereira Soares, S. M., Prystauka, Y., Tat, D., Tomić, A., Voits, T., & Wulff, S. (2023). Monolingual comparative normativity in bilingualism research is out of “control”: Arguments and alternatives. *Applied Psycholinguistics*, 44(3), 316–329.
<https://doi.org/10.1017/S0142716422000315>
- 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>
- Soto-Corominas, A., Daskalaki, E., Paradis, J., Winters-Difani, M., & Janaideh, R. A. (2021). Sources of variation at the onset of bilingualism: The differential effect of input factors, AOA, and cognitive skills on HL Arabic and L2 English syntax. *Journal of Child Language*, 1–33. <https://doi.org/10.1017/S0305000921000246>

- Thornton, R., & Wexler, K. (1999). *Principle B, VP ellipsis, and interpretation in child grammar*. MIT Press.
- Tsimpli, I. M. (2014). Early, late or very late?: Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism*, 4(3), 283–313.
<https://doi.org/10.1075/lab.4.3.01tsi>
- Wang, Y., & Pan, H. (2021). Chinese Reflexives. In Y. Wang & H. Pan, *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.
<https://doi.org/10.1093/acrefore/9780199384655.013.887>
- Woodard, K., Pozzan, L., & Trueswell, J. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology*, 141, 187–209.
<https://doi.org/10.1016/j.jecp.2015.08.005>