

Neuro-cognitive correlates of lexical borrowing during sentence comprehension of bi-dialectal speakers

Junru Wu¹ , Mengru Han¹ and Niels O. Schiller^{2,3,4}

Research Article

Cite this article: Wu, J., Han, M. and Schiller, N.O. (2025). Neuro-cognitive correlates of lexical borrowing during sentence comprehension of bi-dialectal speakers. *Bilingualism: Language and Cognition* 1–14. <https://doi.org/10.1017/S1366728925000264>

Received: 14 May 2024
Revised: 14 February 2025
Accepted: 16 February 2025

Keywords:

lexical borrowing; nativeness; code-switching; bilingualism; speech processing; EEG/ERP

Corresponding author:

Junru Wu;
Email: jrwu@zhwx.ecnu.edu.cn

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

¹Lab of Language Cognition and Evolution, Department Chinese Language and Literature, East China Normal University, Shanghai, China; ²Department of Linguistics and Translation (LT), School of Liberal Arts and Social Sciences (CLASS), City University of Hong Kong, Kowloon, Hong Kong; ³Leiden University Centre for Linguistics, Leiden University, The Netherlands and ⁴Leiden Institute for Brain and Cognition, Leiden University, The Netherlands

Abstract

This study explores lexical borrowing and loanword nativization from a neuro-cognitive perspective testing bi-dialectal speakers of Standard Chinese and Shanghaiese Chinese. We created holistic and morpheme-based cross-dialectal loanwords for auditory sentence processing and compared them with Shanghaiese-specific words, code-switches, and pre-existing etymologically related words. Participants rated their acceptance of each word, indicating Shanghaiese-specific lexical nativeness. GAM analysis of EEG signals revealed that reduced acceptance correlated with frontal positive shifts in ERPs. Holistic loanwords triggered P300-like shifts associated with form-mismatch, whereas morpheme-based loanwords produced LPC-like shifts, suggesting sentence-level re-analysis, and N400-like early frontal negative shifts, indicating lexical integration challenges. Our results indicate that both lexical acceptance and adaptation strategies are pivotal in the cognitive integration of loanwords, revealing distinct neuropsychological stages and pathways in loanword nativization.

Highlights

- Studying lexical borrowing and loan nativization in bi-dialectal speakers using EEG.
- Neuro-cognitive processes of borrowed words in sentence comprehension.
- Reduced acceptance of loanwords with prolonged frontal positive ERP shifts.
- Holistically casted loanwords with P300-like ERP shifts related to form mismatch.
- Morpheme-based loanwords with LPC-like and early frontal negative ERP shifts.

1. Introduction

In a world of constantly evolving linguistic ecology, it is common for languages to coexist and intermix. It becomes inevitable for language users to process a specific sentence structure, wherein a Matrix Language forms the main part, and single words from an Embedded Language are inserted (Myers-Scotton, 2009).

It is important to acknowledge that within this co-evolutionary linguistic ecology, the acceptance of these embedded words as part of the Matrix Language can vary. Furthermore, as language co-evolution progresses, the community's perception of the “nativeness” of these embedded words may shift. With increased acceptance and integration of these embedded words into the Matrix Language, lexical borrowing occurs from the Embedded Language (in this case referred to as the Source Language) into the Matrix language (now referred to as the Recipient Language).

Extensive research demonstrates the significant role of lexical borrowing in language co-evolution, which transforms non-native words and phrases into native ones. Bilingual individuals and those who speak multiple dialects frequently introduce loanwords during their conversations, facilitating the exchange and incorporation of these words between languages. Through this collective behaviour, loanwords gradually become established in the Recipient Language and are perceived as “native” (Giles et al., 1973; Poplack & Sankoff, 1984; Trudgill, 1986). Consequently, borrowing-driven lexical nativization exerts a significant influence on virtually every language worldwide.

Previous studies have investigated how words are adapted cross-linguistically when presented in isolation (Aktürk-Drake, 2014; Ernestus & Baayen, 2003; Kang, 2010; Swerts et al., 2021). However, what neuro-cognitive processes are involved in loanword nativization during sentence comprehension remains largely unclear. To answer this question, it is essential to investigate whether listeners' subjective evaluation of the loanwords' *nativeness* would bias the neuro-

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

cognitive mechanism adopted in comprehension. Also, it is necessary to compare the distinctions between embedded *loanwords* and words with other etymological profiles. These words may include instantaneously inserted *code-switched* single words, translation equivalents that are *etymologically related* and widely accepted as native, as well as words *specific* to the Recipient Language. These comparisons can provide valuable insights into the neuro-cognitive changes that occur during loanword nativization.

To answer these questions, the current study uses electroencephalography (EEG) data to compare lexical forms of various etymological profiles in sentence comprehension. The investigation has been launched into a case involving two dialects so that a broad range of crosslinguistic lexical conditions can be encompassed.

1.1. Typology of embedded words

When considering the “nativeness” of intermixed words embedded into the Recipient Language, it is important to compare loanwords (which have achieved various degrees of acceptance within that language) with other words with different etymological profiles. These may include *code-switched* words (which are not adapted and often still perceived as “foreign” or “non-native” with respect to the Recipient Language), cognates (inherited from the same ancestor language), and words specific to the Recipient Language (e.g., Dixon, 1997; Pagel et al., 2007; Thomason, 2011). Prior to delving into our research questions, it is necessary to review the definitions of these terms. The etymological information for the following example words is sourced from Harper (2001).

- *Code-switched words*: words of one language (Embedded Language, in our case aligned with the Source Language) embedded in sentences of another language (Matrix Language, in our case aligned with the Recipient Language) (Myers-Scotton, 2009), e.g., the Spanish word “pero” used in the English sentence “He came last night *pero* the thing was he stood up Millie’s house” (Zentella, 1997, pp. 99, 119; adapted from Lipski, 2005).¹
- *Loanwords*: words borrowed from one language (Source Language) into another language (Recipient Language) (Crystal, 2003), e.g., the English word “music” used in the English phrase “I like music” originated from Old French “musique”. Note that in an English context with French in the background, it is not guaranteed that an average English speaker would recognize “music” as a loanword. However, it is more likely for English speakers to identify that the word “toufu” in the English phrase “I like toufu” is borrowed.
- *Cognates-a*: words that languages inherited from a shared ancestor (Crystal, 2003), e.g., the English word “father” in the phrase “He is my father” shares a common origin with the Dutch word “vader”, as both originated from Proto-Germanic *fader. However, note that an average English-Dutch bilingual individual may not be aware of this fact.
- *Etymologically Related Translation Equivalent* (ETEs): words of common origin, including both *Cognates-a* (words that languages inherited from a shared ancestor, e.g., English “father” with Dutch in the background) and *Loanwords* (words that languages borrowed from each other, e.g., English “music” with French in the background), also called “related words” by

historical linguists (Crystal, 2003),² but sometimes both are called “cognates” (hence Cognates-b).

- *Recipient-language-specific words*: words specific to the language of the whole sentence (Recipient Language), which were neither borrowed from the language in the background, nor inherited from the same ancestor, e.g., in an English context with French in the background, the English word “child” in the phrase “This is my child” is not related to its French counterpart “enfant” in terms of origins.

Linguistic studies have suggested that etymological origins may influence, yet be independent of, native speakers’ subjective intuition of lexical nativeness (e.g., Poplack & Sankoff, 1984). As previously explained, the term “loanword” encapsulates this dichotomy. While loanwords maintain their etymological uniqueness, their longstanding integration into the everyday lexicon can blur the line between them and cognates in terms of perceived nativeness. Theories of evolutionary linguistics have also highlighted the importance of both etymological origins and loanword nativization in guiding the culture selection that modulates the direction of lexical evolution (e.g., Thomason, 2011). However, further empirical assessment is needed to elucidate the underlying neuro-cognitive basis of these phenomena, which remains largely unexplored. Importantly, the acceptance of specific lexical forms as native to the Recipient Language can significantly differ for loanwords compared to Recipient-Language-specific words and clearly foreign code-switched words. While some loanwords are firmly established and accepted as native, akin to Recipient-Language-specific words, others may continue to be perceived as foreign. These varying acceptance rates of loanwords reflect different social-linguistic stages of loanword nativization. Consequently, these loanwords become ideal candidates for investigating the cognitive processes underlying loanword nativization.

1.2. Lexical nativeness and etymology from a neuro-cognitive perspective

Neuro-cognitive studies on language nativeness often emphasize the distinction between native and non-native listeners (Qiao & Forster, 2017; Weber & Cutler, 2004; Zinszer et al., 2014), focusing on speaker-specific effects in language processing. In contrast, fewer studies have delved into the role of lexical nativeness and etymology in speech comprehension, which pertains to how lexical items are perceived and integrated within a speech community’s native language framework. Our research adds to this discourse by examining the neuro-cognitive aspects of lexical borrowing and loanword nativization, contributing to the ongoing discussion about lexical evolution.

In studies on lexical nativeness, etymologically related translation equivalent (ETE), also referred to as Cognate-b, have been extensively investigated. These include both cognates from common ancestral languages and established loanwords. Bilinguals typically exhibit a ‘cognate effect’ during the cognitive processing of ETEs, showing more efficient processing compared to non-cognate language-specific words (e.g., Dijkstra et al., 2010; Mulder et al., 2015). While fewer studies have examined ETEs in sentences,

¹For more general linguistic reviews of code-switching please refer to Bullock and Toribio’s (2009) and Isurin et al.’s (2009).

²Note that linguistic cognates are not necessarily very similar in their modern forms (e.g., English “mother” and French “mère”) or are exact translation equivalents (e.g., English “image” and Japanese “imeji”, which means impression). However, the current study only focuses on the translation equivalents that are still known as etymologically related.

they report similar advantages (Bultena et al., 2013; Dijkstra et al., 2015). We hypothesize that this efficiency may stem from bilinguals' inclination to perceive ETEs as lexically native to both their languages, thereby potentially enhancing lexical selection. Such processing could underpin the well-supported non-selective activation theory (e.g., Duyck et al., 2007), suggesting that the cognate advantage may be grounded in lexical nativeness. To our knowledge, the cognitive impact of such subjective lexical nativeness evaluations on the processing of ETEs has not been previously explored. Our study investigates this issue.

Another line of research has investigated *single-word code-switching*. Studies of bilinguals' comprehension of code-switching (see Van Hell et al., 2015 for a cognitive review) suggest that it comes with cognitive costs (for reviews of switching cost effects, see Bobb & Wodniecka, 2013; Declerck et al., 2015) both in isolation and in the comprehension of sentences (Bultena et al., 2015; Dijkstra et al., 2015; Guzzardo Tamargo et al., 2016; Kootstra et al., 2012; Litcofsky & Van Hell, 2017; Valdés Kroff et al., 2018; Wang, 2015). ERP studies on cross-linguistic auditory sentence comprehension have indicated two ERP components for code-switching: the N400, which is related to lexical integration, and the Late Positive complex (LPC), related to sentence reanalysis (Van Hell et al., 2015, 2018). Additionally, Liao and Chan's (2016) code-switching research on Mandarin-Taiwanese (Minnan) Chinese bilinguals³ showed evidence of phonological mismatch negativity (PMN) and extensive late negativity. It is unsurprising from a contact-linguistic perspective that switching costs exist, as code-switched forms may be perceived as "foreign" to the Recipient Language. However, for native bilinguals, these forms remain native in the context of their other native language, just like ETEs.

Bringing together the two strands of research, it is evident that embedded ETEs and single-word code-switching tend to have opposite effects on sentence comprehension by native bilingual listeners, although each can be considered to contain native lexical forms for these individuals. Thus, this cognitive contrast may be attributed to the different etymological profiles of their respective lexical forms. Moreover, to date, no cognitive studies have compared different etymological types against one another systematically. However, comprehensively examining the diverse range of etymological profiles, including code-switched forms, loanwords, cognates, and language-specific words, while also taking into account individuals' subjective intuition of lexical nativeness, would facilitate a greater understanding of the cognitive transformations that occur during loanword nativization.

To our knowledge, few neuro-cognitive or psycholinguistic studies have systematically explored the interaction between lexical etymology and individuals' subjective intuition of lexical nativeness. A recent study (Larrazza & Best, 2018) suggested that lexical nativeness may modulate speech comprehension, which compared bilinguals' phonetic-to-lexical mapping strategies to native L1 and non-native L2 lexical forms. However, there is still little known about the influence of language-specific lexical nativeness on proficient early bilinguals.

1.3. Cognitive mechanisms in lexical borrowing

Loanwords are usually phonologically and morphologically adapted to the language they enter (Poplack et al., 1988). This

phenomenon is sometimes referred to as "phonological nativization" (Daniel, 2005; Tarai, 2012). A recent study revealed that native speakers' acceptance of novel words is influenced by formal novelty and lexical regularity (Lombard et al., 2021). Moreover, a recent longitudinal corpus study (Wu & Zhao, 2023) revealed that cross-linguistic adaptation in language co-evolution is not solely based on phonological similarity but is also mediated by systematic phonological correspondence of morphemes, with the complementary relationship between these mechanisms being modulated by linguistic ecology. However, the impact of adaptation on the comprehension of loanwords during the process of nativization remains unclear.

Wu et al. (2021) proposed that there are two basic ways to adapt "foreign" compounds: *holistic casting* and *morpheme-based re-encoding*. (1) *Holistic casting* adapts the loan forms to the Recipient Language's sound system based on similarity and phonotactics. For instance, the Standard-Chinese word 虾米 "dried small shrimps" /ɕiA55mi(214->2)/ can be assimilated into Shanghaiese as /ʃia55mi21/. This process can be taken as reversed from Perceptual Assimilation (e.g., Best & Strange, 1992, PAM). (2) *Morpheme-based re-encoding* involves translating source-language morphemes and combining them into new words in the Recipient Language. For example, the two morphemes that combine to form Standard-Chinese 虾米 /ɕiA55mi2/ "dried small shrimps" can be translated to their etymologically related Shanghaiese morphemes, resulting in the *morpheme-based* loan form /hə53mi44/.⁴ Here, the Shanghaiese-adapted version may not sound very similar to the original two syllables, despite the ontogenesis relationship between the source and the adapted forms. Nevertheless, the morpheme-based re-encoding still results in a loan form that reflects the ontogenesis morphological alignment.

Holistic casting and morpheme-based re-encoding are evident not only in speech production but also in speech comprehension. Bi-dialectals as compared with monolectals are more likely to adopt the morpheme-based mechanism in comprehension borrowing (as well as production borrowing) (Wu et al., 2021). However, bi-dialectals' usage of these mechanisms in comprehension borrowing, especially in sentence context, still requires further investigation.

1.4. Incorporating etymological and acceptance manipulations

The current study investigates the neuro-cognitive signatures associated with the comprehension of lexical borrowing and lexical nativization by proficient and early bi-dialectals of Standard Chinese and Shanghaiese Chinese, using an ERP experiment of sentence comprehension.

To manipulate the etymological profiles of lexical forms, we created two types of paired "nonce" loanwords. For each pair, we compared the loanwords to their corresponding forms in the Recipient Dialect and code-switched forms that convey the same meaning. Furthermore, we incorporated a frequency-matched list of Recipient-Dialect words that have pre-existing etymologically related translation equivalents (ETEs) in the Source Dialect but differ in meaning from the previous types.

Specifically, following Wu et al.'s (2021) theory as reviewed above, the *lexical borrowing* manipulation was split into two conditions: the *morpheme-based* and the *holistically casted*. For instance, for the same concept dried-small-shrimp, the *morpheme-*

³Their case can be taken as Chinese bi-dialectism as well, but involves two more-remotely-related dialects which have much fewer ETEs (according to Tang & Van Heuven, 2009).

⁴Alternatively, a sandhi rule can be applied, yielding /hə44mi21/.

based loanword would be /hə53mi44/ and the *holistically casted* loanword /ʃia55mi21/.

The pair of loanwords was compared against three different references. Firstly, the Shanghaiese-specific compound /khe55fiŋ21/ 开洋 served as the baseline for high etymological nativeness in relation to Shanghaiese.

Secondly, the Standard Chinese source-forms, such as /ɛiA55mi2/ 虾米, acted as the baseline for low etymological nativeness in relation to Shanghaiese. This condition represents code-switching to single words that are considered “foreign” to the Recipient Language.

In addition, pre-existing Shanghaiese translation equivalents (ETEs) were included. These ETEs consisted of well-accepted Shanghaiese compounds that naturally corresponded to their Standard Chinese equivalents and lacked well-known Shanghaiese-specific alternatives. For instance, the Shanghaiese word “friend” /bŋ22iɿ44/ 朋友 aligns etymologically with its Standard Chinese translation equivalent /phŋ35iou214/ 朋友 and lacks notable Shanghaiese-specific alternatives. This comparison aimed to compare between old and nonce borrowings. Also, this condition supports the investigation on the effects of Recipient-Dialect-specific alternatives: the created loanwords have Shanghaiese-specific alternatives, while the pre-existing ETEs have none.

The five etymological conditions are summarized in Table 1. To achieve a better control, the carrier sentences were also made of similar, naturally etymologically related translation-equivalents (ETEs) in Shanghaiese; this etymological profile is the most common lexical context in modern Shanghaiese conversation.

Regarding the comparison between the *loan* forms (1b, 1c in Table 1) and *Shanghaiese-specific* forms (1a in Table 1), *Shanghaiese-specific lexical acceptance* (Recipient-Dialect-specific acceptance) for these forms was also taken into consideration, so that the interaction between etymological nativeness and subjective evaluation of nativeness can be evaluated.

Furthermore, by incorporating Recipient-Dialect-specific acceptance into the design, we can gain a better understanding of the cognitive mechanisms associated with the evolutionary lexical changes during loanword nativization.

Interestingly, we also observed that Shanghaiese-specific forms (1a) displayed varying levels of acceptance as native Shanghaiese words. This indicates a reverse evolutionary direction of lexical change, wherein certain Shanghaiese-specific forms are being replaced by competing loan forms, resulting in reduced acceptance of these Shanghaiese-specific forms among bi-dialectal speakers.

Thus, by measuring Shanghaiese-specific acceptance rates for the three types of forms (1a, 1b & 1c), we aim to compare the neuro-cognitive correlates of the “reduced” lexical acceptance linked to these two types of lexical changes—(lack of) loanword nativization versus obsolescence.

Note that, instead of examining classical bilingual populations of two distinct languages, we tested proficient and early bi-dialectals of Standard Chinese (SC, also known as Putonghua or Mandarin Chinese but in its narrow sense) and Shanghaiese Chinese, who come from bi-dialectal communities in the urban area of Shanghai. As Trudgill (1986) noted, compared to bilingual populations, examining bi-dialectals allows us to gain a more comprehensive view of cross-linguistic dynamics.

The abundance of cross-dialectal lexical phenomena in both past and present motivated us to select Shanghaiese-Standard-Chinese bi-dialectism as the test case. As a Chinese Wu dialect, Shanghaiese has checked tones and lacks diphthong vowels, but is partially overlapping with Standard Chinese (SC) in sound

Table 1. Etymological conditions. Stimuli in Conditions 1a-d organized by concept

	Condition	Explanation	example
(1)	(a) Shanghaiese-specific form	Shanghaiese words that are not etymologically aligned with their SC translation equivalents — high-nativeness baseline	/k ^h ɛ55fiŋ21/ 开洋 “dried small shrimps”
	(b) Morpheme-based loan forms	Loan forms generated by combining Shanghaiese morphemes that are etymologically related to their SC translation equivalents into new Shanghaiese words	/hə53mi44/ 虾米 “dried small shrimps”
	(c) Holistically casted loan forms	Loan forms adapted to the Shanghaiese sound system based on the similarity between the source SC lexical forms and their phonologically most-similar sounds in Shanghaiese, which matches the phonemic inventory and phonotactic regulations of Shanghaiese	/ʃia55mi21/ 虾米 “dried small shrimps”
	(d) Code-switched forms	The source SC lexical forms	/ɛiA55mi2/ 虾米 “dried small shrimps”
(2)	Pre-existing ETEs	Naturally pre-existing Shanghaiese compounds that are etymologically related to their Standard-Chinese equivalents and have no well-known Shanghaiese-specific alternatives	朋友 “friend” / bŋ22iɿ44/ (SC ETE = / p ^h əŋ35iou214/)

inventory and phonotactics (You, 2013). No direct data exists about their mutual intelligibility, but a closely related dialect, Suzhou Wu Chinese, is barely intelligible by SC monolexals (5% in sentence, Tang & Van Heuven, 2009). Although not based on mutual intelligibility, many Shanghaiese words and morphemes are related to their SC counterparts in etymology. This provides us with a baseline of pre-existing ETE words and ETE morphemes to construct morpheme-based loans. There are also pre-existing Shanghaiese-specific words, providing a baseline of high etymological nativeness.

1.5. Research questions and predictions

Drawing on prior research on code-switching-related cognitive costs as reviewed in 1.2, we hypothesize that decreased acceptance of Shanghaiese-specific lexical items will lead to processing challenges similar to those encountered in code-switching. These challenges are expected to manifest in lexical integration and sentence

reanalysis, with corresponding ERP effects such as the N400, LPC, PMN, and extensive late negativity effects. Our analysis of the post-critical-word EEG signal from 0 to 1,500 ms is designed to capture these effects.

Due to the limited availability of Shanghaiese-specific words for sentence composition, our critical words were embedded in sentences with prevalent etymologically related translation equivalents (ETEs), which are the most common elements in the modern Shanghaiese lexicon. This approach differs significantly from previous studies, allowing us to explore new neuro-cognitive territories. Except for the code-switching condition, the other etymological conditions have not been previously investigated neuro-cognitively, prompting us to employ Generalized Additive Modeling to explore timed ERP effects.

Our predictions, which are based on the unique etymological conditions in our study, are as follows: (1) Lexical forms with low acceptance in a native sentence, akin to low cloze probability, may elicit prolonged frontal positivity, as observed by DeLong et al. (2011). (2) Holistically adapted loan forms are anticipated to exhibit stronger form-mismatch effects, such as P300, similar to those reported in code-switching studies (Moreno et al., 2002; Pablos et al., 2019) and in the monitoring of more general cross-linguistic conditions (Von Grebmer zu Wolfsturn et al., 2021) and speech error (Schiller et al., 2009). (3) Morpheme-based loan forms, competing with existing Shanghaiese-specific counterparts, may show enhanced lexical integration effects, indicated by N400-like shifts when compared to naturally occurring ETEs. (4) We anticipate that Shanghaiese-specific forms with disrupted form-meaning connections will trigger the most pronounced effects associated with re-analysis. This expectation is rooted in the difficulty bilingual speakers encounter when attempting to apply their lexical knowledge to decipher these forms. Furthermore, examining these forms will allow us to differentiate the contributions of later lexical-semantic processes, such as the LPC and extended late negativities, from earlier ones, like the N400 (Hubbard & Federmeier, 2021; Kutas & Federmeier, 2011). Despite the loss of overall interpretability, these Shanghaiese-specific forms consist of morphemes that retain meaning and are phonologically recognizable in Shanghaiese. Consequently, we predict that the LPC and/or extended late negativities, which have been associated with reanalysis in previous studies, will be particularly evident in this condition.

These predictions are grounded in the distinct etymological manipulations in our study, which enable us to distinguish the various ERP effects associated with different loan-word conditions. It is important to note that our focus on specific conditions does not preclude the possibility of other ERP effects; rather, it aims to provide a targeted exploration of the anticipated outcomes based on our etymological manipulations.

2. Methods

2.1. Participants

Twenty-six early Standard Chinese-Shanghaiese bi-dialectals were tested (15 female, 11 male, age 18 to 41, $M = 24$, $SD = 6.18$, AOA-Standard Chinese: $M = 2.5$, $SD = 2.70$; AOA-Shanghaiese: $M = 0.54$, $SD = 1.90$ years,⁵ all literate). They were recruited from the urban areas of Shanghai, with internet flyers and word-of-mouth, and received a payment of 100 yuan for their participation

in the 1.5-h experiment. Participants filled in an online questionnaire about their socio-linguistic background by themselves or with the help of a native Shanghaiese speaker to exclude non-urban accents. They reported high proficiency in both dialects (self-rated on a 0-10 scale, SC proficiency: $M = 8.77$, $SD = 1.33$; SH proficiency: $M = 8.46$, $SD = 1.33$).

According to an interview, the bi-dialectal speakers from the urban area of Shanghai “intermix” their two dialects on a daily basis. All participants reported using Standard Chinese and Shanghaiese frequently (self-rated usage-frequencies within the past 3 years on a 0 to 10 scale, SC usage: $M = 9.31$, $SD = 0.97$; SH usage: $M = 6.65$, $SD = 2.59$). We calculated a dialect-dominance index by subtracting SC usage-frequency from Shanghaiese usage-frequency. Twenty participants were Standard Chinese-dominant (dominance index $M = 3.65$, $SD = 2.25$), three were Shanghaiese-dominant (dominance index $M = -1.33$, $SD = 0.58$), and three were balanced. Note that the Shanghaiese-dominant bi-dialectals also use SC very often (frequency rated 8 or 9 out of 10).

2.2. Design and materials

A mixed design was adopted. Etymological conditions of critical words and Shanghaiese-specific acceptance of lexical forms were manipulated.

The stimuli consisted of 56 semantically and grammatically correct sentences. A full listing of the materials is presented in [Supplementary Materials, Appendix 1](#). As designed according to [Table 1](#), the sentence list consisted of two parts. **Part (1)** contained 36 sets of aligned sentences. Each set contained the same Shanghaiese sentence carrying four different types of lexical forms of the same meaning as the critical words: (a) *Shanghaiese-specific forms*, which are not etymologically aligned with their Standard Chinese counterparts, e.g., /k^he55hiɿŋ21/ (开洋) “dried small shrimps”, (b) *morpheme-based loan forms*, which were designed by combining Shanghaiese morphemes that are etymologically related to the critical words’ Standard Chinese translation equivalents, e.g., /hɔ53mi44/ (虾米) “dried small shrimps”, (c) *holistically casted loan forms*, which were designed by assimilating Standard-Chinese words as a whole into Shanghaiese according to the restrictions of Shanghaiese phonology, e.g., /ʃia55mi21/ (虾米) “dried small shrimps”, (d) *code-switched forms*, which are the Standard Chinese versions of the critical words, e.g., /eia55mi214 (→2)/ (虾米) “dried small shrimps”. To form test sets, disyllabic compound nouns specific to the Shanghaiese dialect were selected from Xu and Tao’s (1997) dictionary. These compounds were then assigned *morpheme-based* and *holistically casted* loan forms based on their Standard-Chinese translation equivalents. **Part (2)** contained 20 baseline sentences, which were designed as Shanghaiese sentences with native Shanghaiese critical words that are *pre-existing ETEs* of their Standard-Chinese counterparts and have no well-known Shanghaiese-specific alternatives. In addition, nine native Shanghaiese sentences were included for practice at the beginning but not included in the analysis.

All the sentences consisted of five words and seven-to-eight syllables, and the position of the critical word within each sentence was always located at the third word and after the same disyllabic Shanghaiese phrase 伊讲 /i55 gaŋ13/ ‘3rd-Person SAY’. Each sentence contained a two-word trisyllabic phrase following the critical word, which is in Shanghaiese, e.g., 味道好 /mi22dɔ44 fiɔ13/ ‘tastes good’. All critical words were disyllabic (34 in Shanghaiese, 33 in SC) or trisyllabic (two in Shanghaiese, three in SC) nouns. The four types of lexical variants aligned for the critical

⁵We made our best efforts to control the Age of Acquisition (AOA) within a small range, but it was not incorporated into the modeling process.

Table 2. Durations of sentences, mean durations & onset latencies for critical words (in ms, *SD in brackets*)

	durations of sentences	onset latencies for critical words	durations of critical words
Shanghainese-specific form	1455.42 (114.02)	349.33(45.94)	404.89(76.31)
Morpheme-based loan forms	1519.38 (115.44)	383.78(38.56)	421.86(66.93)
Holistically casted loan forms	1520.59 (143.15)	385.06(57.73)	446.22(106.22)
Code-switched forms	1456.30 (110.16)	356.25(40.05)	424.11(66.62)
Pre-existing ETes	1578.45(106.53)	386.15(37.95)	434.75(62.08)

words were counterbalanced across participants, who nevertheless listened to all the Part-(2) baseline sentences. Thus, concepts of the critical words were not repeated for the same participant. An early and highly proficient female Standard Chinese-Shanghainese bi-dialectal recorded the sentences at 44,100 Hz. She had PSC level 2a for Standard Chinese and was verified with extensive Shanghainese lexical knowledge by a questionnaire. Comparing her Shanghainese accent to the documentation (You, 2013) revealed that she has a typical urban young-generation accent with some middle-generation features.

The recorded sentences on average lasted 1,523 ms (*SD* = 118). The critical words begin on average at 369 ms (*SD* = 47) after the onset of the sentences and lasted 426 ms (*SD* = 79). As shown in Table 2, the natural differences regarding duration and the onset latency of critical words across conditions did not bias our analyses.

2.3. Procedure and EEG recording

Participants participated in two tasks: (1) an auditory judgment task according to information provided by the stimuli sentences with EEG data recorded, and (2) a Shanghainese sound-image verification task.

2.3.1. (1) EEG experiment

Participants were instructed in Shanghainese that they would listen to separate sentences, after hearing each sentence they would hear a question and see a picture on the screen; then they had to decide whether the image was a proper answer to the question according to the information given in the sentence and press keys on a keyboard to indicate their choices. The question and picture appeared 300 ms after the end of each test sentence.

Participants were seated 50 to 55 cm away from the screen in a sound-attenuated electromagnetic shielding chamber. The EEG data were recorded with 64 Ag/AgCl electrodes mounted in an elastic cap (NeuroScan Inc., USA). Electrode locations consisted of eight sites along the midline (FPZ, FZ, FCZ, CZ, CPZ, PZ, POZ, OZ), 27 left lateral electrodes (FP1, AF3, F1, F3, F5, F7, FC1, FC3, FC5, FC7, C1, C3, C5, T7, CP1, CP3, CP5, TP7, P1, P3, P5, P7, PO3, PO5, PO7, O1, CB1), and 27 right lateral electrodes (FP2, AF4, F2, F4, F6, F8, FC2, FC4, FC6, FC8, C2, C4, C6, T8, CP2, CP4, CP6, TP8, P2, P4, P6, P8, PO4, PO6, PO8, O2, CB2). Bipolar recordings were made above and below the left eye to monitor for vertical eye movements and blinks. Bipolar recordings of the outer canthus of the right and left eyes were made to monitor for horizontal eye

movements. Electrodes were re-referenced off-line to the average of the left and right mastoids (M1 & M2). A NeuroScan SynampsRT amplifier amplified the electroencephalogram (EEG) at a sample rate of 1,000 Hz. Electrode impedances were kept below 10kΩ.

2.3.2. (2) Sound-image verification

Shanghainese-specific lexical acceptance of critical words was collected with a Shanghainese-specific sound-image verification task in E-Prime 3.0 (Schneider et al., 2002) after the EEG experiment. The *loan lexical forms* and *Recipient-Dialect-specific lexical forms* (see Table 1 and Section 2.2 Part-1, abc) were measured with a *Shanghainese-specific* sound-image verification task to assess the corresponding *Recipient-Dialect-specific acceptance* for each lexical form, as an indication of the general subjective evaluation within the linguistic community regarding the nativeness of each lexical form.

The participants were aurally instructed in Shanghainese that they would see objects and hear Shanghainese, and they were required to press a key in each trial to indicate whether the sound they heard was the way they named the object or the Shanghainese people around them named the object in Shanghainese. This task was conducted after the EEG experiment so that their judgement would not influence the processing of stimuli sentences. To avoid confusion from code-mixing during the judgement only critical lexical forms for the Part-1(a)(b)(c) conditions were rated. The proportion of acceptance responses across participants was calculated for each critical lexical form in order to model it further. Figure 1 displays the by-concept distribution of *Shanghainese-specific lexical acceptance*, with the three dimensions representing the mean nativeness rated for the *Shanghainese-specific form*, the *morpheme-based loan form*, and the *holistically casted loan form* associated with the same concept.

2.4. Analysis of EEG data

2.4.1. (1) Preprocessing

An off-line 0.1 Hz high-pass filter was applied first to remove big and slow body-motion artifacts. Then Independent Component Analysis (ICA) from EEGLAB2019.0 (Delorme & Makeig, 2004) was applied to each participant's recording, to remove blinking, vertical eye-movement, horizontal eye-movement, muscle-related high-frequency artifacts, and remaining motion artifacts. Only components that can be attributed to clear causes of artifacts were removed, and any artifact components that may be mixed with brain-related activities were kept. The artifact-free data went through an off-line 30 Hz low-pass filter.

Onsets of critical words within stimulus sentences were manually marked and extracted with Praat (Boersma & Weenink, 2017) scripts before the experiment. We used a MATLAB (2019) script to mark the codes for lexical forms at the onsets of the critical words within the corresponding participants' EEG data. EEG signal curves were extracted between -200 and 1,500 ms post-critical word onset with a baseline of 200 ms pre-critical-word activity, referring to previous studies on the comprehension of code-switching (following Fernandez et al., 2019; Litcofsky & Van Hell, 2017).

Trials with any remaining artifacts were excluded from the analyses, first automatically with ERPLAB (Lopez-Calderon & Luck, 2014) using a 100 dB threshold and then manually. Percentages of excluded trials for each condition were: 5.13% for Shanghainese-specific forms (12/234), 2.54% for morpheme-based loan forms (6/236), 3.45% for holistic-casted loan forms (8/232),

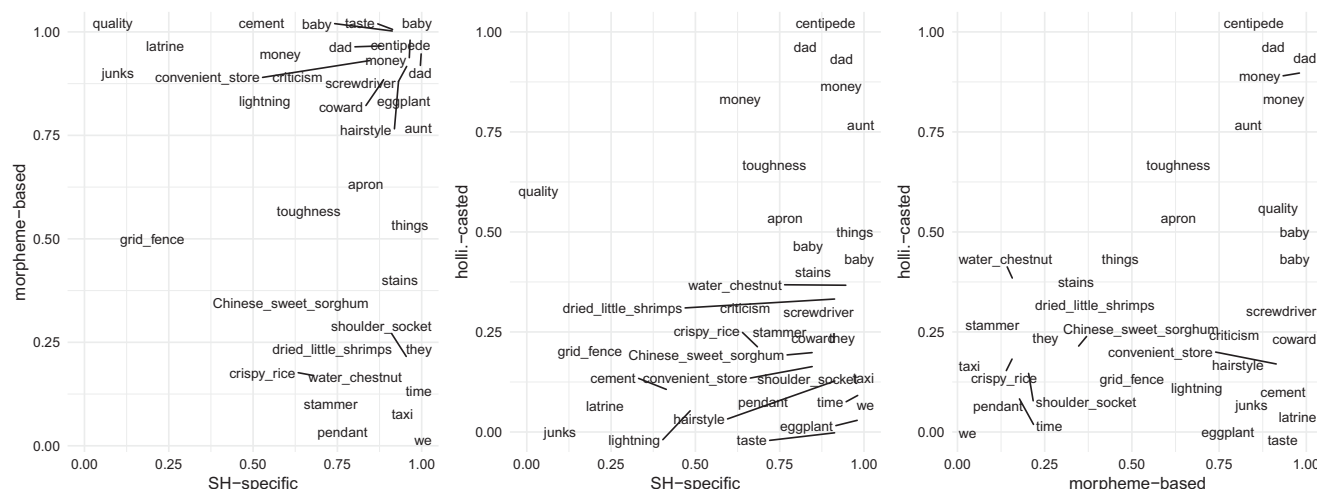


Figure 1. Two-dimensional planes illustrating the Shanghainese-specific acceptance of lexical forms (0-1) for Shanghainese-specific forms (SH-specific), morpheme-based loan forms (morpheme-based), and holistically casted loan forms (holli. casted). Concepts are indicated by English text labels.

2.14% for code-switching (5/234), and 1.54% for pre-existing ETEs (baseline, 8/520).

2.4.2. (2) GAM analysis

The EEG signal was time-locked to the onset of the critical words (Shanghainese-specific forms, morpheme-based loan forms, holistically casted loan forms, code-switched to SC forms, and the Shanghainese version of pre-existing ETEs).

A counterbalanced design was implemented to prevent the repetition of synonyms within the same participant's experience. Consequently, each participant completed nine separate trials for each of the conditions (a, b, c, d) and participated in all 20 trials for the pre-existing ETE condition (condition 2), which serves as a baseline. This design decision, coupled with the limited availability of suitable Shanghainese-specific words, resulted in a smaller number of critical lexical forms per participant compared to what might be used in a classical ERP study. Despite our efforts to control for factors such as length and frequency across different types of critical forms, there remained potential lexical differences, such as the distribution of phonemes, which could complicate the interpretation of ERP effects.

To address these challenges and to make the most of the data we collected, we employed Generalized Additive Modelling (GAM). GAM allows us to directly model the non-linear dynamics of EEG signals over time, capturing the detailed interaction between these signals and the Shanghainese-specific acceptance of lexical forms (Abugaber et al., 2023). A significant advantage of GAM is its capacity to model the effects from specific lexical forms and to account for idiosyncratic participant effects on EEG signals in a non-linear manner. These effects are incorporated into the model with random smooths, which effectively separate them from the main experimental effects. This approach not only mitigates the potential confounds of lexical variation and lower trial numbers but also provides a more detailed and sensitive analysis than traditional ERP methods, particularly suited to our data and research questions.

Thus, no EEG signal waveforms were collapsed across trials. Moreover, with GAM, we did not set time windows of interest either because precise time windows with significant differences can be and were detected post-hoc from the model results. Nevertheless, EEG signals were still collapsed within regions of interest

(ROI) factors involving anteriority (anterior, posterior) and laterality (right, left hemisphere). Following Litcofsky and Van Hell (2017) and Fernandez et al. (2019), electrodes were grouped into four ROIs: left frontal ("LF": F3, F5, F7, FC1, FC3, FC5); right frontal ("RF": F4, F6, F8, FC2, FC4, FC6); right posterior ("RP": CP2, CP4, CP6, P4, P6, P8), and left posterior ("LP": CP1, CP3, CP5, P3, P5, P7). We also modeled EEG signals from three mid-line electrodes (Fz, Cz, Pz). Considering the limitation of computational power⁶, separate models were built for each ROI and mid-line electrode.

We used the 'mgcv' package (Wood, 2017) in R (R Core Team, 2019) to perform GAM and significantly different time windows across conditions were identified with time-sensitive post-hoc analyses on the GAM models, which were conducted with the 'plot_diff' function from 'itsadug' R package (Van Rij et al., 2019) using a 1.96 criterion for standard error.

Three sets of GAM analyses were conducted with EEG amplitude as the dependent variable: (1) The first model included a three-level categorical predictor *etymological* condition to compare *morpheme-based loan forms* and *holistically casted loan forms* against *Shanghainese-specific forms* (baseline). The non-linear effect of *time* and its non-linear interaction with z-scaled Shanghainese-specific nativeness of lexical forms (referred to as *nativeness* in the tables) were modelled within the same smooth function using the "s" type of spline. The participant- and item-induced variations were included in the random smooths. We built candidate models in a backward-dominant way (see the [Supplementary Materials, Rmd codes](#) for details). These models were compared base on the Akaike Information Criterion likelihood values (Sakamoto & Ishiguro, 1986). (2) The second model also included a three-level categorical predictor *etymological* condition to compare the two types of *loan forms* against *code-switched forms* (baseline). Since no data of Shanghainese-specific nativeness of lexical forms were acquired for code-switched lexical forms (which are just SC forms), only the non-linear effect of *time* on EEG amplitude were assessed with GAMs using the "s" type of spline. The participant- and item-induced variations were included in the random smooths in a similar way as in the first set. We also

⁶The final models reported in the current study was fit in Ubuntu Xenial 16.8 environment with a SWAP of 40 GB & intel core i7, taking around 20 h.

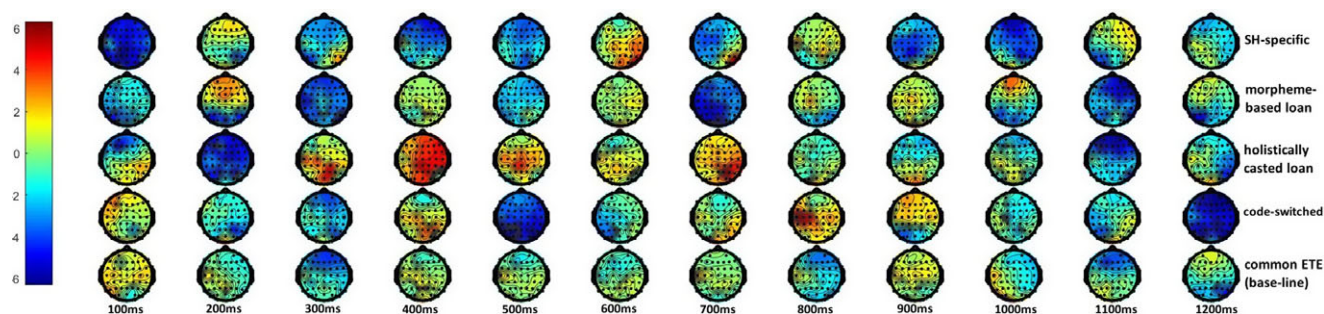


Figure 2. Topographic maps. Rows from top to bottom: (1) Shanghainese-specific forms, (2) morpheme-based loan forms, (3) holistically casted loan forms, (4) code-switched to SC, (5) pre-existing ETEs without Shanghainese-specific alternatives. Positivity is colored in red and negativity in blue.

further stratified the data of the two types of loan forms into parts with high (scaled nativeness ratings ≥ 0.5), median (scaled nativeness ratings between -1.1 and 0.5), and low ratings of nativeness (scaled nativeness ratings ≤ -1.1) and built separate models to compare them with the code-switched forms. (3) The third model was built similarly to the models in (2) but compared the two types of *loan forms* against *pre-existing ETEs*.

3. Results: ERP effects modelled with GAM

Participants had a mean accuracy of 95.6% ($SD = 3.2$) on judgement questions during the EEG recording, which indicates that they paid attention to the stimuli.

See Figure 2 for the topographic maps of brain activity.

3.1. GAM analyses comparing loan forms against Shanghainese-specific forms

The ultimate arrangement of the first GAM model that compares the two types of *loan forms* with the *Shanghainese-specific forms* is displayed in the highest cell of [Supplementary Materials, Appendix 2 \(Table 3\)](#). The two random terms modelled the by-participant and by-item random smooths of time, indicating the non-linear influences from participant individuals and specific lexical forms.

T-statistics for the parametric coefficients compared the two loan conditions against the baseline. The parametric coefficients showed that morpheme-based loan forms elicited a significant positive main effect as compared with Shanghainese-specific forms, in the left posterior region, $t = 3.79$, $p < 0.001$, as well as at FZ, $t = 2.47$, $p < 0.05$, CZ, $t = 2.00$, $p < 0.05$. F-statistics for the smooth terms compared each manipulation level with the average level. Hence to answer the research questions, post-hoc analysis is necessary. We created surface plots for the partial effects of model estimates (Figure 3) and ERP-like sliced curve plots for estimated partial effects (Figure 5, first rows in each panel). We also calculated the estimated difference between smooths to generate estimated differences (Figures 4 & 5) for ERP effects (Van Rij et al., 2019).

Firstly, taking ERP responses to Shanghainese-specific forms that were mostly accepted as Shanghainese-native (scaled acceptance = 1.1) as the baseline for nativeness, the estimated difference waveforms compared the other combinations of etymological conditions (three levels) and scaled acceptance rates (high = 1.1, median = -0.1 , low = -1.5) against it. Time chunks with significant differences in ERP waveforms in this comparison were marked with dash-separated time chunks with red solid lines in-between along

the horizontal axes in Figure 4. Taking together the shapes of difference waveforms, ERP effects were generally more positive with a decrease in acceptance as well as with the use of *loan forms*, both exhibiting an increase in positivity. There were likely two positive components, one before 600 ms, the other after 600 ms following critical words.

Regarding time and scalp distributions, the reduction of nativeness was largely responsible for the increase of early positivity, although it had a varying effect on different Regions of Interest (ROI) under different etymological conditions. For instance, in addition to eliciting frontal positivity, an early ERP response to *Shanghainese-specific forms* (shown in red in Figure 4) was more pronounced in the right posterior region: when the acceptance rate was at a median level, the positive shift began at 92 ms, whereas with low acceptance it started at 6 ms; by comparison, in the left posterior region this shift began much later, at 727 ms with median acceptance and 796 ms with low acceptance. Moreover, *Shanghainese-specific forms* with low acceptance elicited the largest positive shift in ERP within all the non-baseline conditions in the majority of regions and electrodes (except for the left posterior region)

Similarly, regarding *morpheme-based loan forms* (shown in green in Figure 4), the lowest acceptance rates, in addition to eliciting frontal positivity, primarily increased right posterior positivity significantly within a 40–1,294 ms timeframe after the critical words. This was much earlier than in the left posterior region, where this shift began at 693 ms post-critical words.

Comparatively, *holistically casted loan forms* (shown in blue in Figure 4) are more likely to elicit more positive-going timeframes before 600 ms post-critical words. Moreover, early ERP responses to *holistically casted loan forms* with high acceptance rates (left lower sub-plots in the panels of Figure 4) were not only more positively going than forms with median and low acceptance rates, but also showed greater activation in the left and frontal sites, while yielding no significant timeframes in the right posterior region.

The post-hoc comparison between EEG signals of the two types of *loan forms* with each other, as well as against the *Shanghainese-specific forms* was conducted, with scaled acceptance rates fixed to high ($=1.1$), median ($=-0.1$), and low ($=-1.5$). The sliced partial effect waveforms and difference waveforms are shown in Figure 5.

Given the high acceptance rates (shown in left sub-plots in each panel of Figure 5), both types of loan forms elicited positive shifts of ERP responses after 600 ms as compared with Shanghainese-specific forms. However, in earlier time windows (<600 ms), *morpheme-based loan forms* (b, green) hardly elicited more-positively going timeframes compared to Shanghainese-specific

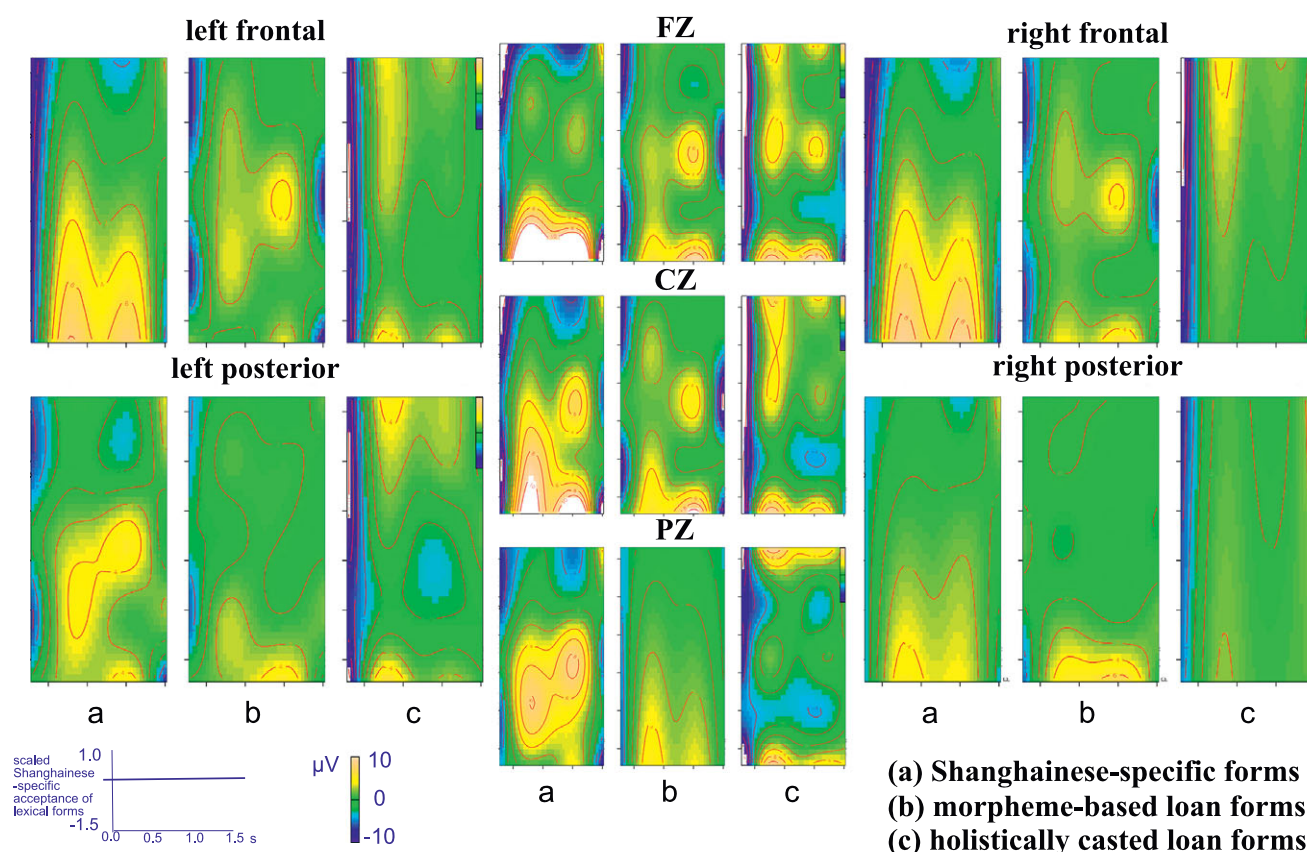


Figure 3. The bi-dialectals' partial effects for the interaction of time (horizontal axis) and scaled Shanghaiese-specific acceptance of lexical forms (vertical axis), in left frontal (upper left), left posterior (lower left), right frontal (upper right), right posterior (lower right), electrode FZ (upper middle), electrode CZ (center), and electrode PZ (lower middle). Sub-plots within each panel represent the three etymological conditions: (a) Shanghaiese-specific forms (left), (b) morpheme-based loan forms (middle), (c) holistically casted loan forms (right). In each panel, the x-axis shows time in seconds, the y-axis displays scaled acceptance of Shanghaiese-specific lexical forms, and color denotes partial GAM smooth effects on EEG signals.

forms (a, red), except at the FZ electrode, where the positive time-frame started at 29 ms. In contrast, *holistically casted loan forms* (c, blue) elicited more-positively going timeframes that started very early (earlier than 10 ms), compared to both Shanghaiese-specific forms (a, red) and morpheme-based loan forms (green, b-a; purple, c-b). With regard to scalp distributions, the positive ERP shifts for morpheme-based loan forms were right-dominant, while those for holistically casted loan forms were left-dominant (see significance marker in Figure 5 for the rest of the details).

Given the median acceptance rates (shown in the middle sub-plots in each panel of Figure 5), the ERP differences across the three etymological conditions were more subtle. When compared to *Shanghaiese-specific forms* (b-a, green), *morpheme-based loan forms*, elicited less-positively going timeframes late in the left-posterior region (951 to 1,122 ms) and at the PZ electrode (848 to 1,119 ms), whereas *holistically casted loan forms* (c, blue) elicited a more-positively going early timeframe at the FZ electrode (212 to 487 ms) and a late timeframe at PZ electrode (762 to 1,157 ms) (c-a, blue). Furthermore, *holistically casted loan forms* (c, blue) also elicited more-positively going early timeframes in left posterior (127 to 436 ms) and right posterior (195 to 521 ms) regions (c-b, purple) when compared to morpheme-based loan forms (b, green).

Given low acceptance rates, the *Shanghaiese-specific forms* (a, red) elicited the largest positive ERP shifts involving both early and later timeframes, especially in the frontal regions, while the two types of loan forms elicited less positively going shifts (b-a, green &

c-a, blue). Moreover, *holistically casted loan forms* (c, blue) elicited a more positively going shift in an early timeframe (229 to 435 ms) in the left frontal region, as compared with morpheme-based loan forms (b, green) (c-b, purple). These results are shown in right sub-plots in each panel of Figure 5, with significance markers for details.

In summary, the results of this section indicate that: (1) both borrowing and reduced nativeness ratings are associated with positive ERP shifts. (2) *Morpheme-based loan forms* primarily lead to late (>600 ms post critical words) positive ERP shifts, predominantly involving the right hemisphere. (3) *Holistically casted loan forms* are related to early (before 300 ms post critical words) as well as late positive ERP shifts, with the early shifts greater. (4) The most pronounced ERP shifts in both early and late time windows were observed for *Shanghaiese-specific forms* with low acceptance rates and *holistically casted loan forms* with high acceptance rates, yet these ERP effects were right-dominant and left-dominant, respectively, in scalp distribution.

3.2. GAM analyses comparing loan forms against code-switched forms

The ultimate arrangement of the second GAM model that compares the two types of *loan forms* with the *code-switched forms* was: $\text{amp} \sim s(\text{time}, \text{by} = \text{etymology}) + \text{etymology} + s(\text{time}, \text{participantid}, \text{bs} = \text{"fs"}, m = 1) + s(\text{time}, \text{item}, \text{bs} = \text{"fs"}, m = 1)$.

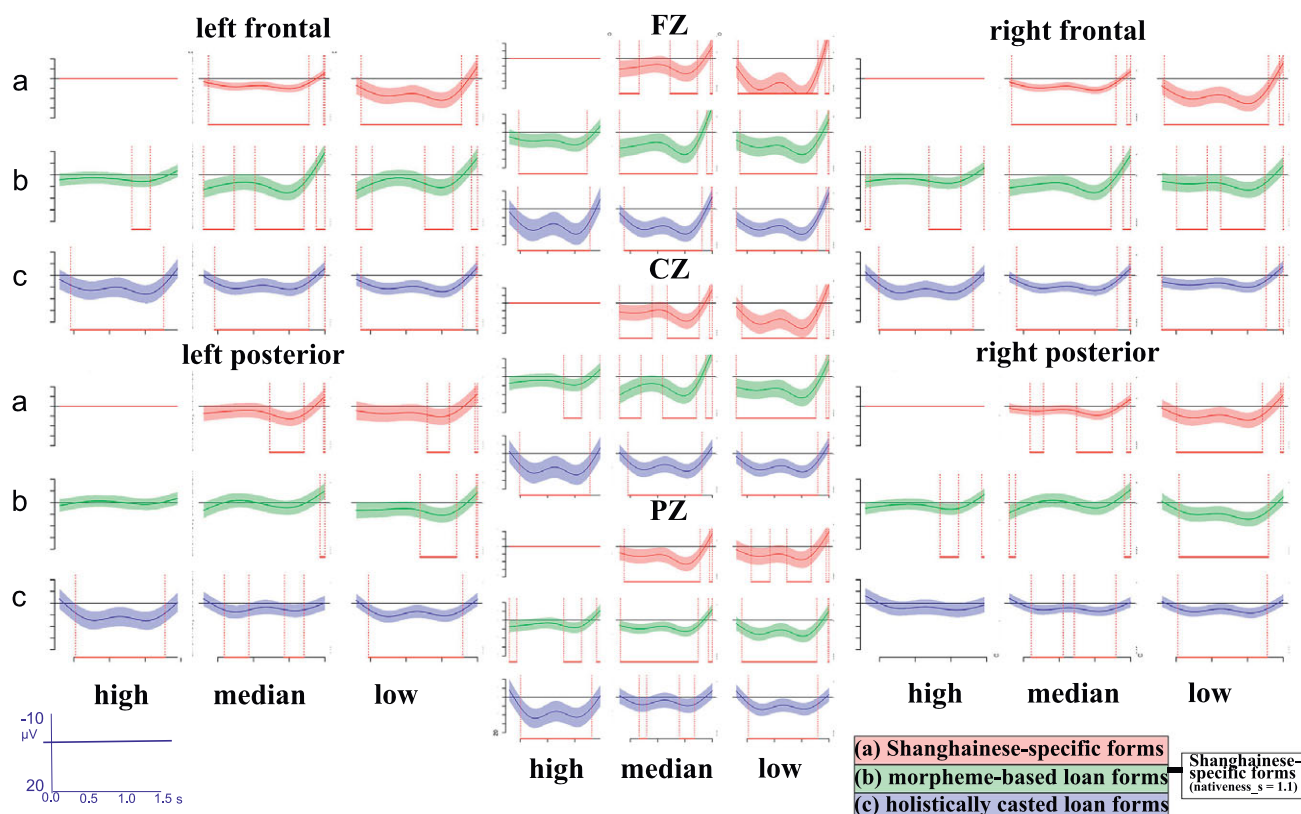


Figure 4. Estimated difference waveforms comparing (a) *Shanghaiese-specific forms* (red, first row in each panel), (b) *morpheme-based loan forms* (green, second row in each panel), and (c) *holistically casted loan forms* (blue, third row in each panel), with high (left column in each panel), median (middle column in each panel), and low (right column in each panel) acceptance rates as Shanghaiese-native, against Shanghaiese-specific forms with highest acceptance rates (high-nativeness baseline), in left frontal (upper left), left posterior (lower left), right frontal (upper right), right posterior (lower right), electrode FZ (upper middle), electrode CZ (centre), and electrode PZ (lower middle), significant time chunks marked.

Here we report *t*-statistics for significant parametric terms and significant time-frames calculated from the estimated difference between smoothes (Van Rij et al., 2019). Full model statistics are reported in the tables in [Supplementary Materials, Appendix 2 \(Table 4\)](#).

When compared to *code-switched forms*, *holistically casted loan forms* elicited significant positive shifts in the left posterior region, $t = 2.36$, $p < 0.05$. The post-hoc difference estimates were significant in a 23 to 126 ms post-critical-word timeframe on the ERP waveforms in this ROI. This positive left-posterior ERP difference, as shown by the stratified models, also applied for loan forms with high acceptance rates, $t = 3.27$, $p < 0.01$, significant timeframes = -114^7 to 143 ms, 332 to 847 ms, and 1,071 to 1,277 ms, as well as for loan forms with low acceptance rates, $t = 3.09$, $p < 0.01$, significant timeframes = 6 to 212 ms, 487 to 624 ms, 762 to 951 ms, and 1,242 to 1,483 ms. In addition, *holistically casted loan forms* with low acceptance rates also elicited significant positive ERP shifts in other sites, including the left frontal region, $t = 2.20$, $p < 0.05$, the right posterior region, $t = 2.00$, $p < 0.05$, significant timeframe = 1,380 to 1,500 ms, FZ, $t = 2.05$, $p < 0.05$, significant timeframe = 1,311 to 1,380 ms, and CZ, $t = 2.50$, $p < 0.05$, significant timeframes = 40 to 178 ms and 916 to 933 ms.

⁷The mean of the 200 ms pre-critical-word activity was used as the baseline reference when we extracted ERP waveforms. However, here the time chunk of positive shift still started 114 ms before the onset of the critical word.

Neither holistically casted loan forms with median ratings of nativeness, nor morpheme-based loan forms showed any significant difference when compared to code-switched forms.

3.3. GAM analyses comparing loan forms against pre-existing ETEs

The ultimate arrangement of the second GAM model that compares the two types of *loan forms* with the *pre-existing ETEs* and the full model statistics are also reported in [Supplementary Materials, Appendix 2 \(Table 5\)](#). Here we only report *t*-statistics for significant parametric terms and significant timeframes calculated from the estimated difference between smoothes (Van Rij et al., 2019).

When compared to *pre-existing ETEs*, *morpheme-based loan forms* elicited a significant negative ERP shift in the left frontal region, $t = -3.30$, $p < 0.001$. The post-hoc difference estimates were significant in an early timeframe (-63 to 659 ms post critical word) in this ROI. Additionally, stratified models revealed that this early negative left-frontal ERP shift also applied for morpheme-based loan forms with high acceptance rates, $t = -3.11$, $p < 0.001$, significant timeframe = -114 to 659 ms. Moreover, *morpheme-based loan forms* with high acceptance rates also elicited significant negative ERP shifts in other sites, that is, early in the right frontal, $t = -3.00$, $p < 0.001$, significant timeframe = 10 to 624 ms, late in the left posterior, $t = -2.10$, $p < 0.05$, significant timeframe = 1,260 to 1,311 ms, and right posterior, $t = -2.07$, $p < 0.05$ (long-lasting difference but with no significant post-hoc time-frames).

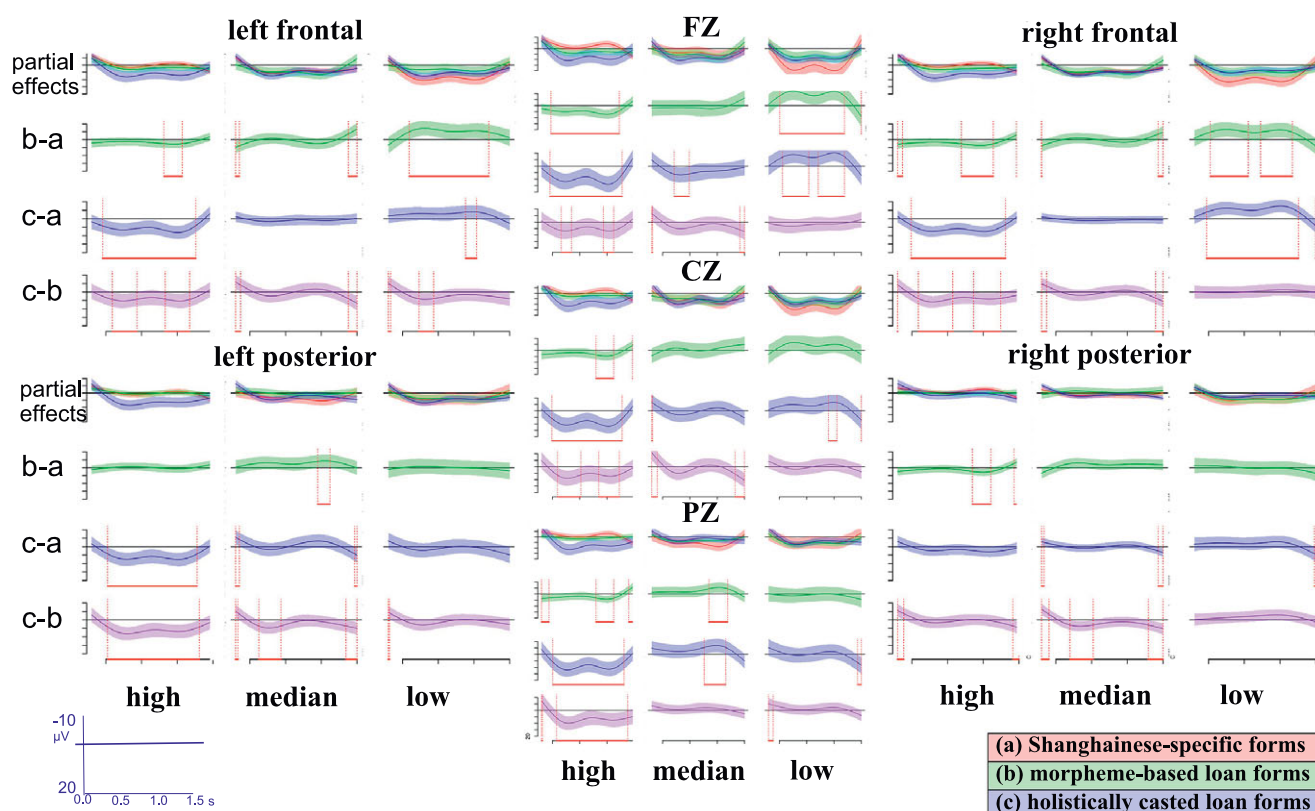


Figure 5. Partial effects (first row in each panel) of (a) Shanghaiese-specific forms (red), (b) morpheme-based loan forms (green), and (c) holistically casted loan forms (blue), as well as estimated difference waveforms comparing these three conditions against each other (b-a, green, second row; c-a, blue, third row; c-b, purple, fourth row in each panel), with high (left column in each panel), median (middle column in each panel), and low (right column in each panel) acceptance rates, in left frontal (upper left), left posterior (lower left), right frontal (upper right), right posterior (lower right), electrode FZ (upper middle), electrode CZ (centre), and electrode PZ (lower middle), significant timeframes marked.

Neither holistically casted loan forms, nor morpheme-based loan forms with median or low ratings of nativeness showed any significant difference as compared with pre-existing ETE forms.

4. Discussion

In this study, we sought to investigate the neuro-cognitive basis for lexical borrowing and loanword nativization in sentence comprehension. Using EEG measurements, we investigated bi-dialectals' speech comprehension of two types of loanwords (holistically casted and morpheme-based), compared to dialect-specific, code-switched, and pre-existing etymologically related translation equivalent forms. Additionally, we investigated the influence of the Recipient-Dialect-specific acceptance of lexical forms as an indicator for the influence of subjective lexical nativeness.

4.1. Prolonged frontal positivity, etymology, and the reduction of acceptance

A general finding of this study suggests that, in comparison to highly accepted Recipient Dialect-specific words (high-nativeness baseline), all other conditions elicited a prolonged positive shift in frontal ERPs during auditory sentence comprehension. As not consistently identified in previous neuro-cognitive studies on bilinguals, this shift may be attributed to the combined influence of etymology and acceptance (the result of nativization).

Firstly, the EEG differences observed may be attributed to etymological origins rather than participants' intuition of lexical nativeness, as loan forms systematically differ from highly native Shanghaiese-specific words only in terms of etymology, not in acceptance within the Recipient-Dialect. Yet, this finding may still be grounded in the participants' knowledge—bi-dialectal participants are aware that the loanwords accepted in the Recipient Dialect have corresponding forms in the Source Dialect and hence can deduce that these words are not part of the original vocabulary of the Recipient Dialect.

Secondly, we saw that the decreased acceptance of obsolete Recipient-Dialect-specific lexical forms also introduced similar positive frontal ERP shifts [see the red (a) waveforms in Figure 4]. Since these words originate from the Recipient Language itself, this finding indicates that frontal positive ERP shifts are also related to an independent effect of the loss of lexical nativeness.

However, the effects of nativization on loan forms may be more intricate, as revealed by these ERP effects. As compared with the high-nativeness baseline, the two types of loan forms, regardless of their level of Recipient-Dialect-specific acceptance, elicited prolonged frontal positive shifts of ERPs [see the green (b) and blue (c) waveforms in Figure 4]. Nevertheless, for each etymological type, the reduced acceptance tends to show non-linear effects [see the (b) and (c) panels in Figure 3]. Furthermore, a comparison between loan forms and Recipient-Dialect-specific forms of similar acceptance [see the (b-a) and (c-a) waveforms in Figure 5], also revealed intricate influences of etymology.

Overall, these results suggest that etymology and acceptance of loan forms may both influence processing, though they interact in a non-linear way. It is likely that the reduction of language-specific acceptance for a given lexical form and having etymological knowledge about its external origin both can trigger real-time ERP effects, which reflect violations of listeners' expectations for lexical forms, in agreement with previous findings related to the reduction of cloze probability (DeLong *et al.*, 2011). Thus, in terms of evolution, fine-tuning to bi-dialectal lexical expectations in native sentences may be the key neuro-cognitive step for loanwords to be accepted as native.

4.2. LPC-like effects and sentence-level reanalysis

Examining Figures 4 and 5 in more detail, we found that loan forms elicited LPC-like late positive shifts when compared to Recipient-Dialect-specific forms. This suggests that borrowing can also trigger sentence-level reanalysis, which was also found by both Liao and Chan's (2016) and Fernandez *et al.* (2019) in their studies on intrasentential code-switching.

Moreover, compared with code-switched forms, holistically casted loan forms with high or low acceptance rates evoked larger late positive ERP shifts in the left posterior region (see Section 3.2). This implies that the adaptation of the form of holistically casted borrowing requires a stronger sentence-level reanalysis, possibly also related to instantaneous online creation of new lexical representations (Wu *et al.*, 2021) in the left temporal lobe.

This cognitive barrier indicated by the LPC-like effects may explain why only a small fraction of new or nonce loanwords tend to remain in the Recipient Languages in the documented social linguistic cases (Poplack *et al.*, 1988; You, 2016).

4.3. Early left frontal positivity and form mismatch

In Sections 4.1 and 4.2, we have demonstrated that the ERP effects of holistically casted and morpheme-based loanwords are very similar. However, they also triggered neuro-cognitive differences. Particularly, holistically casted loan forms elicited greater early positive ERP shifts, especially in the left and frontal sites and especially for those with high acceptance rates (see Figures 3, 4, and 5). Additionally, the same kind of differences were observed when comparing holistically casted loan forms with code-switched forms (see Section 3.2). This effect overlaps with the P300 ERP component, which was proposed to reflect match/mismatch with working memory trace and be sensitive to the probability of stimulus (Patel & Azzam, 2005). A P300 was also found in code-switching experiments, though in visual mode (Moreno *et al.*, 2002). Based on the current findings, it is possible that holistic casting, which includes phonological adaptation and the creation of mental representations for phonologically novel forms, leads to a more pronounced form mismatch in the bi-dialectals' working memory. Consequently, this increased discrepancy captures greater attention during the early stages of processing. In terms of evolution, variations in cognitive loads may result in an uneven selection process among different loan-form alternatives.

4.4. Early left frontal negativity and morpheme-based re-encoding

Although N400 effects had been found in most studies on code-switching and had been associated with lexical integration (e.g., Fernandez *et al.*, 2019), surprisingly, we did not find N400-like

effects in most of the comparisons. This suggests that lexical integration was not the main issue introduced by cross-dialectal borrowing and that the *cognate advantage* may be more important (Bultena *et al.*, 2013) for the cognitive processing of loanwords.

However, we did observe early negative ERP shifts around 400 ms (although starting quite early) in the left frontal region when comparing morpheme-based loan forms against pre-existing ETEs (see Section 3.3). Moreover, stratified modelling showed that this difference remained true for highly accepted morpheme-based loan forms, which are also ETEs in nature. In this case, the ERP difference could only be attributed to the fact that the highly native morpheme-based loanwords have Recipient-Dialect-specific alternatives, which may introduce stronger lexical-level competition and affected lexical integration, explaining the difference. Thus, from an evolutionary standpoint, bi-dialectals and bilinguals may be less likely to adopt new lexical forms to denote concepts that have already been represented in a Recipient Language in order to avoid the increased cognitive costs associated with lexical competition, conforming to a Primacy Effect (Sam, 2013). However, notably previous research (Wu *et al.*, 2021) showed that bi-dialectals do not necessarily adhere to the mutual exclusivity principle (Markman, 1992) when comprehending loan forms.

4.5. Neuro-cognitive correlates of obsolescence

Although still known by our two young Shanghaiese informants, some Shanghaiese-specific lexical forms were rated with low acceptance rates in the Recipient Dialect Shanghaiese (e.g., /leʔ1seʔ13/ "grid fence"). Obviously, these forms are becoming obsolete. As noted in Section 4.1, the reduced acceptance of these Shanghaiese-specific lexical forms is related to prolonged positive shifts of ERPs. In fact, the obsolete Shanghaiese-specific lexical forms elicited the largest positive frontal ERP shifts among all the conditions.

Importantly, the obsolete Shanghaiese-specific lexical forms however differ from the other less accepted forms in the scalp distribution of ERP effects. For instance, they did not elicit early ERP shifts in the left posterior region but instead triggered early and late right posterior positivity.

Note that these obsolete lexical forms etymologically originate in the Recipient Language, so that they may still be phonemically and morphologically familiar. The successful phonemic and morphemic retrieval may explain the lack of left posterior effects. Difficulties in processing these obsolete forms seem to be more of a semantic nature. Form-meaning associations of these words have already been broken, while they are not similar enough to their SC translation equivalents for the bi-dialectals to guess the corresponding lexical meanings. This explains the right posterior positivity and large late frontal positivity responses. As we have associated prolonged positivity with the reduction of nativeness in the previous discussion, we can claim here that, neuro-cognitively speaking, obsolete Recipient Dialect-specific forms are even less "native" than the loan forms least accepted by the bi-dialectal community, which matches our research prediction (4). The observed ERP effects and the subsequent disruption of form-meaning links potentially contribute to the cognitive basis for both individual language loss and the extinction of languages at the community level.

In our analysis, we employed Generalized Additive Modeling (GAM) to dissect the EEG data, revealing cognitive processing dynamics akin to traditional ERP components but without predefined time windows. GAM's capacity to model non-linear relationships and accommodate participant variability made it an optimal

choice for our study's design, despite the limited trials per condition. This approach provides a robust platform for exploring cognitive and neural mechanisms in language processing, as evidenced by the ERP-like effects observed.

Taken together, this study showed that lexical borrowing and loanword nativization are associated with bi-dialectal individuals' cognitive processing. Such processing involves both phonological and lexical challenges, but the amount of attentional resources and difficulty of reanalysis may vary depending on the linguistic adaptation employed for borrowing. The findings of this study offer insight into the cognitive basis for the selection of cross-dialectal loan forms in co-evolving dialects and languages.

Supplementary material. To view supplementary material for this article, please visit <http://doi.org/10.1017/S1366728925000264>.

Acknowledgments. We would like to thank the LCE Lab's research assistants Zhexuan Zhang, Wei Chen, Yaxin Li, Yiqi Zheng, and Junyuan Zhao for their meticulous efforts in recruiting suitable participants, and especially Yiqi Zheng and Junyuan Zhao as our Shanghaiese informers for their time and insights. This work was supported by the Chinese Fundamental Research Funds for the Central Universities (2017ECNU-YYJ017), and by the Shanghai Philosophy and Social Sciences Fund (2017BY001).

Data availability statement. The data that support the findings of this study are openly available in Open Science Framework at <https://osf.io/ny5td/>, reference number DOI 10.17605/OSF.IO/NY5TD.

Ethical standard. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Competing interest. The authors have no competing interests to declare.

References

- Abugaber, D., Finestrat, I., Luque, A., & Morgan-Short, K. (2023). Generalized additive mixed modeling of EEG supports dual-route accounts of morphosyntax in suggesting no word frequency effects on processing of regular grammatical forms. *Journal of Neurolinguistics*, *67*, 101137. <https://doi.org/10.1016/j.jneuroling.2023.101137>.
- Aktürk-Drake, M. (2014). The role of perceptual salience in bilingual speakers' integration of illicit long segments in loanwords. *Lingua*, *143*, 162–186. <https://doi.org/10.1016/j.lingua.2014.02.006>.
- Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*, 305–330.
- Bobb, S. C., & Wodniecka, Z. (2013). Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology*, *25*, 568–585.
- Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [Computer program] Version 6.0.3 retrieved 2017 from <http://www.praat.org/> [Computer software].
- Bullock, B. E., & Toribio, A. J. (2009). Themes in the study of code-switching. In B. E. Bullock & A. J. Toribio (Eds.), *The Cambridge Handbook of Linguistic Code-switching*. Cambridge University Press.
- Bultena, S., Dijkstra, T., & van Hell, J. G. (2013). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1214–1241. <https://doi.org/10.1080/17470218.2013.853090>.
- Bultena, S., Dijkstra, T., & Van Hell, J. G. (2015). Language switch costs in sentence comprehension depend on language dominance: Evidence from self-paced reading. *Bilingualism: Language and Cognition*, *18*(03), 1–17. <https://doi.org/10.1017/S1366728914000145>.
- Crystal, D. (2003). *A dictionary of linguistics & phonetics*. Wiley-Blackwell.
- Daniel, O. (2005). *Phonological nativization of Dholuo loanwords* [PhD Thesis], University of Pretoria. <http://hdl.handle.net/2263/27396>.
- Declerck, M., Koch, I., & Philipp, A. M. (2015). The minimum requirements of language control: Evidence from sequential predictability effects in language switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 377–394. <https://doi.org/10.1037/xlm0000021>.
- Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations: Dual ERPs to low probability sentence continuations. *Psychophysiology*, *48*(9), 1203–1207. <https://doi.org/10.1111/j.1469-8986.2011.01199.x>.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.
- Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*(3), 284–301. <https://doi.org/10.1016/j.jml.2009.12.003>.
- Dijkstra, T., Van Hell, J. G., & Brenders, P. (2015). Sentence context effects in bilingual word recognition: Cognate status, sentence language, and semantic constraint. *Bilingualism: Language and Cognition*, *18*(4), 597–613. <https://doi.org/10.1017/S1366728914000388>.
- Dixon, R. M. W. (1997). *The rise and fall of languages*. Cambridge University Press.
- Duyck, W., Assche, E. V., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 663. <https://doi.org/10.1037/0278-7393.33.4.663>.
- Ernestus, M., & Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, *79*(1), 5–38.
- Fernandez, C. B., Litcofsky, K. A., & van Hell, J. G. (2019). Neural correlates of intra-sentential code-switching in the auditory modality. *Journal of Neurolinguistics*, *51*, 17–41. <https://doi.org/10.1016/j.jneuroling.2018.10.004>.
- Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*, *2*(2), 177–192. <https://doi.org/10.1017/S0047404500000701>.
- Guzzardo Tamargo, R. E., Valdés Kroff, J. R., & Dussias, P. E. (2016). Examining the relationship between comprehension and production processes in code-switched language. *Journal of Memory and Language*, *89*, 138–161. <https://doi.org/10.1016/j.jml.2015.12.002>.
- Harper, D. (2001). *Online etymology dictionary*. Online Etymology Dictionary.
- Hubbard, R. J., & Federmeier, K. D. (2021). Dividing attention influences contextual facilitation and revision during language comprehension. *Brain Research*, *1764*, 147466. <https://doi.org/10.1016/j.brainres.2021.147466>.
- Isurin, L., Winford, D., & de Bot, K. (2009). *Multidisciplinary Approaches to Code Switching* (Vol. 41). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.41>.
- Kang, Y. (2010). The emergence of phonological adaptation from phonetic adaptation: English loanwords in Korean. *Phonology*, *27*(2), 225–253. <https://doi.org/10.1017/S0952675710000114>.
- Kootstra, G. J., Van Hell, J. G., & Dijkstra, T. (2012). Priming of code-switches in sentences: The role of lexical repetition, cognates, and language proficiency. *Bilingualism: Language and Cognition*, *15*(04), 797–819. <https://doi.org/10.1017/S136672891100068X>.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>.
- Larraza, S., & Best, C. T. (2018). Differences in phonetic-to-lexical perceptual mapping of L1 and L2 regional accents. *Bilingualism: Language and Cognition*, *21*(4), 805–825. <https://doi.org/10.1017/S1366728917000323>.
- Liao, C.-H., & Chan, S.-H. (2016). Direction matters: Event-related brain potentials reflect extra processing costs in switching from the dominant to the less dominant language. *Journal of Neurolinguistics*, *40*, 79–97. <https://doi.org/10.1016/j.jneuroling.2016.06.004>.
- Lipski, J. M. (2005). Codeswitching or Borrowing? No sé so no puedo decir, you know. In L. Sayahi & M. Westmoreland (Eds.), *Selected Proceedings of the Second Workshop on Spanish Sociolinguistics* (pp. 1–15). Cascadilla Proceedings Project.

- Litcofsky, K. A., & Van Hell, J. G. (2017). Switching direction affects switching costs: Behavioral, ERP and time-frequency analyses of intra-sentential code-switching. *Neuropsychologia*, *97*, 112–139. <https://doi.org/10.1016/j.neuropsychologia.2017.02.002>.
- Lombard, A., Huyghe, R., & Gyga, P. (2021). Neological intuition in French: A study of formal novelty and lexical regularity as predictors. *Lingua*, *254*, 103055. <https://doi.org/10.1016/j.lingua.2021.103055>.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*. <https://doi.org/10.3389/fnhum.2014.00213>.
- Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In *Modularity and constraints in language and cognition* (Vol. 25, pp. 59–101). Lawrence Erlbaum Associates.
- MATLAB. (2019). Version 9.6.01072779 (R2019a) [Computer software]. The MathWorks Inc.
- Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching *Palabras* (Words): An electrophysiological study of code switching. *Brain and Language*, *80*, 188–207. <https://doi.org/10.1006/brln.2001.2588>.
- Mulder, K., Brekelmans, G., & Ernestus, M. (2015). *The processing of schwa reduced cognates and noncognates in non-native listeners of English*. In *The 18th International Congress of Phonetic Sciences [ICPhS 2015]*.
- Myers-Scotton, C. (2009). *Duelling languages: Grammatical structure in codeswitching* (Repr.). Clarendon.
- Pablos, L., Parafita Couto, M. C., Boutonnet, B., De Jong, A., Perquin, M., De Haan, A., & Schiller, N. O. (2019). Adjective-noun order in Papiamentu-Dutch code-switching. *Linguistic Approaches to Bilingualism*, *9*(4–5), 710–735. <https://doi.org/10.1075/lab.17036.pab>.
- Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*(7163), 717–720. <https://doi.org/10.1038/nature06176>.
- Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected studies of the event-related Potential. *International Journal of Medical Sciences*, *147*–154. <https://doi.org/10.7150/ijms.2.147>.
- Poplack, S., & Sankoff, D. (1984). Borrowing: The synchrony of integration. *Linguistics*, *22*(1). <https://doi.org/10.1515/ling.1984.22.1.99>.
- Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, *26*, 47–104.
- Qiao, X., & Forster, K. I. (2017). Is the L2 lexicon different from the L1 lexicon? Evidence from novel word lexicalization. *Cognition*, *158*, 147–152. <https://doi.org/10.1016/j.cognition.2016.10.026>.
- R Core Team. (2019). R: a language and environment for statistical computing [Computer program]. R Foundation for Statistical Computing, Vienna, Austria, version 3.6.2 [Computer software].
- Sakamoto, Y., & Ishiguro, M. (1986). *Akaike information criterion statistics*. Reidel.
- Sam, N. M. S. (2013). Primacy effect. *PsychologyDictionary.Org*. <https://psychologydictionary.org/primacy-effect/>.
- Schiller, N. O., Horemans, I., Ganushchak, L., & Koester, D. (2009). Event-related brain potentials during the monitoring of speech errors. *NeuroImage*, *44*(2), 520–530. <https://doi.org/10.1016/j.neuroimage.2008.09.019>.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Incorporated.
- Swerts, M., Van Heteren, A., Nieuwendorp, C., Von Oerthel, E., & Kloots, H. (2021). Asymmetric Forms of Linguistic Adaptation in Interactions Between Flemish and Dutch Speakers. *Frontiers in Communication*, *6*, 716444. <https://doi.org/10.3389/fcomm.2021.716444>.
- Tang, C., & Van Heuven, V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*, *119*(5), 709–732. <https://doi.org/10.1016/j.lingua.2008.10.001>.
- Tarai, S. (2012). Phonological nativization of Arabic, Portuguese and English loanwords in Odia. *Journal of Portuguese Linguistics*, *11*(2), 79. <https://doi.org/10.5334/jpl.87>.
- Thomason, S. G. (2011). *Language contact: An introduction* (Repr.). Edinburgh Univ. Press.
- Trudgill, P. (1986). *Dialects in contact*. B. Blackwell.
- Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Dussias, P. E. (2018). Experimental contributions of eye-tracking to the understanding of comprehension processes while hearing and reading code-switches. *Linguistic Approaches to Bilingualism*, *8*(1), 98–133. <https://doi.org/10.1075/lab.16011.val>.
- Van Hell, J. G., Fernandez, C. B., Kootstra, G. J., Litcofsky, K. A., & Ting, C. Y. (2018). Electrophysiological and experimental-behavioral approaches to the study of intra-sentential code-switching. *Linguistic Approaches to Bilingualism*, *8*(1), 134–161. <https://doi.org/10.1075/lab.16010.van>.
- Van Hell, J. G., Litcofsky, K. A., & Ting, C. Y. (2015). Intra-sentential code-switching: Cognitive and neural approaches. In J. W. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 459–482). Cambridge University Press. <https://doi.org/10.1017/CBO9781107447257.020>.
- Van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2019). Itsadug: Interpreting Time Series, Autocorrelated Data Using GAMMs. [Computer program] Retrieved from <https://cran.r-project.org/web/packages/itsadug> [Computer software].
- Von Grebmer zu Wolfsturn, S., Pablos, L., & Schiller, N. O. (2021). Noun-phrase production as a window to language selection: An ERP study. *Neuropsychologia*, *162*, 108055. <https://doi.org/10.1016/j.neuropsychologia.2021.108055>.
- Wang, X. (2015). Language control in bilingual language comprehension: Evidence from the maze task. *Frontiers in Psychology*, *6*. <https://doi.org/10.3389/fpsyg.2015.01179>.
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. [https://doi.org/10.1016/S0749-596X\(03\)00105-0](https://doi.org/10.1016/S0749-596X(03)00105-0).
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC.
- Wu, J., & Zhao, J. (2023). Systematic correspondence in co-evolving languages. *Humanities and Social Sciences Communications*, *10*(1), 469. <https://doi.org/10.1057/s41599-023-01975-6>.
- Wu, J., Zheng, W., Han, M., & Schiller, N. O. (2021). Cross-dialectal novel word learning and borrowing. *Frontiers in Psychology*, *12*, 734527. <https://doi.org/10.3389/fpsyg.2021.734527>.
- Xu, B., & Tao, H. (1997). *Shanghai Fangyan Cidian [Shanghai Dialect Dictionary]*. Jiangsu Education Press.
- You, R. (Ed.). (2013). *Shànghǎi dìqū fāngyán diàochá yánjiū* (上海地区方言调查研究. A linguistic survey of Shanghai dialects) (1st ed.). Fudan Daxue Chubanshe (复旦大学出版社, Fudan University Press).
- You, R. (2016). Words of Foreign Origin in the 'Shanghai Tongshu Ji Yang-jingbang' (《上海通俗语及洋泾浜》所见外来词研究). In *Dialect contact treatise* (方言接触论稿) (pp. 153–180). Fudan University Press.
- Zentella, A. C. (1997). *Growing up bilingual: Puerto Rican children in New York*. Blackwell Publishers.
- Zinszer, B. D., Malt, B. C., Ameel, E., & Li, P. (2014). Native-likeness in second language lexical categorization reflects individual language history and linguistic community norms. *Language Sciences*, *5*, 1203.