Short title: Shanghainese vowel merger reversal

Full title: On the cognitive basis of contact-induced sound change: Vowel merger reversal in Shanghainese

Names, affiliations, and email addresses of authors:
Yao Yao
The Hong Kong Polytechnic University
ctyaoyao@polyu.edu.hk

Charles B. Chang
Boston University
cc@bu.edu

Address for correspondence:
Yao Yao
The Hong Kong Polytechnic University
Department of Chinese and Bilingual Studies
Hung Hom, Kowloon, Hong Kong

Short title: Shanghainese vowel merger reversal

Full title: On the cognitive basis of contact-induced sound change: Vowel merger reversal in Shanghainese

Abstract

This study investigated the source and status of a recent sound change in Shanghainese (Wu, Sinitic) that has been attributed to language contact with Mandarin. The change involves two vowels, /e/ and /ɛ/, reported to be merged three decades ago but produced distinctly in contemporary Shanghainese. Results of two production experiments showed that speaker age, language mode (monolingual Shanghainese vs. bilingual Shanghainese-Mandarin), and crosslinguistic phonological similarity all influenced the production of these vowels. These findings provide evidence for language contact as a linguistic means of merger reversal and are consistent with the view that contact phenomena originate from cross-language interaction within the bilingual mind.*

Keywords: merger reversal, language contact, bilingual processing, phonological similarity, crosslinguistic influence, Shanghainese, Mandarin.

---

* This research was supported by funding from the Dean's Reserve for Research, Scholarly, and Other Endeavors, Faculty of Humanities, The Hong Kong Polytechnic University (project number: 1-ZV8U). The authors are grateful to Yan Jiang, Rong Li, Chang Liu, Chun-ling Catherine Liu, Jia Lou, Weiwen Xu, and Shengru Yao for research assistance and to Danny Erker, Brook Hefright, Gregory Iverson, the Associate Editor, two anonymous reviewers, and audience members at the CUHK International Conference on Bilingualism and Comparative Linguistics, the 13th Conference on Laboratory Phonology, and the 20th Annual Conference of the International Association of Chinese Linguistics for helpful comments.

1. INTRODUCTION. What happens when two languages come into contact? This question is explored in both second language acquisition (SLA) research and socio-historical linguistics, but from different perspectives: whereas SLA studies focus on the individual, socio-historical studies consider the speech community as a whole. These core emphases have led to distinct research programs related to the analysis and explanation of language contact phenomena. A fundamental issue in SLA is the interaction between the first and second languages (L1 and L2, respectively) of an L2 learner, which may or may not lead to language change in the speech community. On the other hand, socio-historical studies aim to understand the social context of interaction between language groups and the transmission and diffusion of ensuing linguistic change.

The division of labor between SLA and socio-historical linguistics has led to the investigation of language contact at both 'micro' (INDIVIDUAL) and 'macro' (COMMUNITY) levels; however, it is also at odds with the widely held belief that 'the locus of language contact is the bilingual speaker' (Sankoff 2002:643, see also Weinreich 1968). To put it another way, consideration of the individual is inseparable from the study of contact-induced language change. For sound change in particular, Sapir (1921:213) observed that 'we may suppose that individual variations arising at linguistic borderlands—whether by the unconscious suggestive influence of foreign speech or the actual transfer of foreign sounds into the speech of bilingual individuals—have been gradually incorporated into the phonetic drift of a language'. That is, macro change (in the language of a speech community) starts with micro change (in the idiolect of a member of that community).

The enterprise of linking community-level change with individual-level processing and production—evident in much of the early research related to bilingualism (e.g. Weinreich 1968)—has not generally translated to research in SLA and socio-historical linguistics as the two fields have developed increasingly divergent concerns. To be sure, the relevance of individual-level factors, such as aspects of language experience (e.g. language dominance, proficiency, usage frequency, usage patterns, etc.) and differences in social integration, is acknowledged in studies of language contact (Lai & Hsu 2013, Nagy 1997), but not discussed in detail. The language contact literature is also biased toward the examination of crosslinguistic influence (or IMPOSITION) from a dominant language to a secondary or less proficient language (e.g. Guy 1990, Ratte 2011, van Coetsem 1995, Winford 2005), which raises the question of the extent to which previously documented patterns of contact influence are limited to certain contact situations (i.e. those characterized by a marked asymmetry in dominance between the languages in contact) or to a certain direction of influence (i.e. from the dominant language/L1 to the secondary language/L2).

In the current study, we aim to revitalize the enterprise of connecting COMMUNITY with INDIVIDUAL in investigations of language change. In addition to reflecting the inherent linkage between the language of the group and the idiolect of the individual, we argue that examining the speech community and the individual in tandem benefits research in both fields. Recent developments in models of SLA and bilingualism help account for the linguistic outcomes of language contact observed in a speech community; language contact, in turn, provides an opportunity to test these models with different types of bilinguals, as patterns of language use often vary across generations and/or other social groups in contact situations.

Our test bed for linking COMMUNITY with INDIVIDUAL was an ongoing sound change in contemporary Shanghainese that is occurring allegedly due to contact with Mandarin. The change involves two vowels, /e/ and /ɛ/, which were once nearly merged but are currently becoming distinct again. Aside from providing an interesting example of merger reversal, this case provided an opportunity to examine crosslinguistic influence in the less frequently studied L2-to-L1 direction. So we conducted a study of this sound change with two specific goals: (a) providing a linguistic account of the ongoing change, situated in its social and historical context, and (b) probing the individual speaker's internal mechanisms of language processing and production that have ultimately led to the change. Thus, it should be noted that what we mean by INDIVIDUAL is the cognitive basis—within an individual mind—of contact phenomena observed at the community level, not the concept of 'individual differences' central to much SLA research. In short, our aim is to demonstrate how consideration of the specific dynamics of language interaction within the bilingual mind can inform the prediction and explanation of contact-induced language change.

The rest of the paper is organized as follows. Section 2 summarizes the literature on merger reversals, introduces the socio-historical background of the sound change under investigation, and lays out the theoretical framework adopted in this paper. Section 3 describes the experimental methods of the current study. Section 4 presents the acoustic data obtained from production experiments and reviews the results of relevant statistical analyses. Finally, Section 5 discusses the implications of our findings for models of bilingual speech and theories of contact-induced sound change, followed by a brief conclusion in Section 6.

2. BACKGROUND.

2.1. REVERSING A MERGER.

Merger of phonemic categories—a reductive diachronic change—is a historical development commonly observed across languages, whereas the reversal of such a merger is rare. Since a merger reversal effectively recreates a contrast where there was none, it has long been claimed that a true reversal of merger—that is, an 'unassisted' reversal of a complete merger—is, in principle, impossible. According to GARDE'S PRINCIPLE (Garde 1961), a complete merger cannot be reversed by linguistic means. Thus, previously documented cases of merger reversal are regarded as exceptional, in that they involve a merger that was not truly complete and/or a means of reversal that was not truly linguistic.

Most examples of merger reversal involve reversal of an incomplete merger. In these cases, the relevant contrast never fully disappeared at the phonetic level, with the sounds at issue continuing to show subtle, yet reliable, differences within an individual speaker or for some sector of the relevant speech community (Labov 1975, 1994, Labov et al. 1991, cf. Baranowski 2007); hence, merger in these cases is typically described as NEAR-MERGER. A salient feature of near-merger is an asymmetry between perception and production: speakers preserve a subtle distinction in production, but fail to distinguish the sounds in perception. Crucially, the preservation of a distinction in production allows for a full resurrection of the contrast—in both perception and production—at a later time point. Labov and colleagues documented a few cases of near-merger in progress (Labov et al. 1991) and proposed that the famous *meat-mate* merger reversal in the history of English was due to the near-merger (as opposed to complete merger) of the *meat* and *mate* word classes (Labov 1975).

Other examples of merger reversal may be explained as 'assisted' reversal (for a review, see Trask 2000:286–287). In some cases, the relevant contrast was neutralized phonetically, but left a trace of its former existence in different phonological roles of the merged segments, which enabled merged lexical items to be separated into their original lexical sets; this is what happened in the case of the /j/-/ʃ/ merger reversal in Gipuzkoan Basque (Michelena 1957). In other cases, the reestablishment of the previous contrast was aided by the fact that the membership of merged words in different lexical sets was systematically encoded in the language's orthography (e.g. Kochetov 2006). In yet other cases, the contrast was simply borrowed from a different variety of the language maintaining the contrast, typically for sociolinguistic reasons related to prestige (Ihalainen 1994, Weinreich et al. 1968).

Although previously documented cases of merger reversal differ in the details of what was reversed and how the reversal was accomplished, they have in common a return to a linguistic state of affairs that is similar to a prior stage of the language. The operative word here is 'similar', however, because it is not clear for any of these cases that the newly contrastive variants emerging from the merger reversal were phonetically IDENTICAL to the originally contrastive variants that preceded the merger. On the contrary, given the amount of time that typically passes between the pre-merger and post-reversal stages of a language that has undergone a merger reversal (on the order of a century), newly contrastive variants are likely to differ in one or more respects from their pre-merger counterparts. Merger reversal should, therefore, imply phonetic similarity, but not necessarily phonetic identity, between original and innovative variants; in fact, in certain cases (e.g. [j] > [ʃ] > [χ] in Gipuzkoan Basque; Michelena 1957) the innovative variant is quite different from the original form. As such, the crucial part of this phenomenon is recovery of contrast (a contrast that may or may not be identical in phonetic realization to the original contrast), and it is in this sense that we will be using the term MERGER REVERSAL in this paper.

The merger reversal in Shanghainese examined here is noteworthy in the context of the literature on merger reversals because, as discussed in more detail below, it is not amenable to an explanation in terms of any of the aforementioned exceptions to Garde's principle. Although the merger of Shanghainese /e/ and /ɛ/ may not have been complete prior to its reversal, the innovative realization of this contrast belies the original, monophthongal quality of both phonemes and, therefore, cannot be a mere replication of remnants of the prior contrast (see §2.2). Furthermore, this merger reversal could not have been assisted in the manner of previously documented merger reversals because the contrast is not cued by phonological restrictions or distinct spellings and is absent (and, thus, not borrowable) from the contact language (i.e. Mandarin). The Shanghainese case thus presents a significant challenge to our current understanding of merger reversals, and below we provide some background on the socio-historical context in which this change occurred.

2.2. SOCIO-HISTORICAL CONTEXT OF THE SHANGHAINESE MERGER REVERSAL.

A BRIEF HISTORY OF THE SHANGHAINESE LANGUAGE. A member of the Wu family of Chinese languages,[1] Shanghainese is spoken mainly in the municipal area of Shanghai, China. Like other Wu dialects, spoken Shanghainese is mutually unintelligible with the country's standard language, Mandarin, but the two languages share many cognate words. Literate speakers of Shanghainese mostly use Standard Written Mandarin as their written language, although the brief history of writing in Shanghainese left a legacy of writing conventions for some native Shanghainese words, which are not unfamiliar to today's speakers.

Over the past 150 years, spoken Shanghainese was brought into contact with other Chinese languages via two social processes: immigration and implementation of national language policy. The modern history of Shanghai dates back to the mid-19th century, when the city became one of the first Chinese treaty ports open to foreign ships. In the following 100 years, the city experienced a tremendous population increase (Chen 1995, Qian 2007), due in large part to a vast wave of immigration from the nearby Jiangsu and Zhejiang provinces. Early immigrants and their children quickly adopted Shanghainese as their main language, while also introducing into the language features from their home languages. As a result, Shanghainese in this period was heavily influenced by surrounding Wu dialects, especially those of Suzhou and Ningbo.

Ever since the 1950s, when Mandarin (or Putonghua) started being promoted as the standard language of China, Shanghainese has been in increasing contact with Mandarin as a consequence of the country's language policy. Standard Mandarin (both spoken and written) became almost ubiquitous in various forms of mass media. It also became the official language of instruction in educational institutions, although some schools got away with using Shanghainese for non-language subjects in the first couple of decades after promotion of Mandarin began in earnest. In the late 1980s, however, according to the younger participants in our study, the Putonghua education policy was reinforced, with most urban schools teaching all subjects in Mandarin and requiring students to speak Mandarin both inside and outside of class. Speaking

Mandarin was avidly promoted as an important aspect of a civilized life. Consequently, for individuals born after 1980, Mandarin became the preferred language for discussing formal topics, and code-switching between Shanghainese and Mandarin became common practice. Over the past twenty years, Mandarin has thus overtaken nearby Wu dialects in becoming the most influential contact language for Shanghainese. Scholars have noticed many Mandarin features being transferred into Shanghainese, such that the local language spoken by young people in Shanghai today carries a strongly perceptible Mandarin accent (e.g. Qian 2003, 2007).

The sound change examined in this paper manifests both strata of contact influence discussed above: the earlier influence from surrounding Wu dialects and the more recent influence from Mandarin. In the next section, we discuss the historical development of the sound change in more detail.

MID FRONT VOWELS IN SHANGHAINESE. Contemporary Shanghainese is described as having an inventory of nine vowel phonemes, with a total of 14 vowel qualities that are distributed complementarily between open- and closed-syllable environments (Chen 2008). The lack of phonological alternations in the language makes it difficult to tell how the closed-syllable vowels [ɪ ʏ ʊ ə ɐ] should be identified with the open-syllable vowels [i y ø ɛ o u ɤ ɔ a], but this phonemic ambiguity is unimportant for the purposes of this study. Of relevance to the mid front vowels of interest is the fact that there are also two mid back vowels /o/ and /ɔ/, which can appear in the same open-syllable environments and are thus unambiguously contrastive, as well as a mid front rounded vowel /ø/.[2]

The sound change under investigation concerns two mid front unrounded vowels, close-mid /e/ and open-mid /ɛ/. These two vowels were contrastive phonemes in Shanghainese in the mid-19th century (Edkins 1868); were merged or nearly merged in the 1980s due to influence from other Wu dialects (e.g. Xu & Tang 1988); and have recently been reported to be distinct again, ostensibly due to Mandarin influence (e.g. Gu 2007). The current paper focuses on the most recent change from (near-)merger to recovered distinctiveness.

Lexical items that participated in the series of changes involving /e/ and /ɛ/ belong to three lexical sets, which can be distinguished by the rhymes in their Mandarin counterparts: [aj], [an], and [ej]. For clarity, we will refer to these lexical sets as MN-[aj] items, MN-[an] items, and

MN-[ej] items, respectively.[3] Table 1 summarizes four stages in the historical development of the rhymes in these lexical sets.[4] In Stage I, both MN-[aj] and MN-[ej] items were pronounced with [e] and only MN-[an] items were pronounced with [ɛ] (Edkins 1868). In Stage II, MN-[aj] items changed to [ɛ] (Chao 1928, Karlgren 1926), but the distinction between /e/ and /ɛ/ was still preserved, as MN-[ej] items maintained the [e] pronunciation. In Stage III, MN-[ej] items changed to [ɛ] as well, resulting in a total merger of /e/ and /ɛ/. As Chen (1995) observed, the source of this merger was most likely influence from the Suzhou dialect, which pronounced all three lexical sets with [ɛ] consistently, as attested in both the 1920s and 1990s.

<INSERT TABLE 1>

Although /e/ and /ɛ/ were ostensibly merged in Stage III, the status of the merger is highly debatable. Currently available documentation suggests a wide range of possibilities, from complete or near-complete merger to total separation. Based on fieldwork conducted in urban districts in the late 1970s and early 1980s, Xu and Tang (1988) claimed that the merging of /e/ and /ɛ/ was the prevailing trend among middle-aged and younger speakers at that time. Therefore, only /ɛ/ was listed in the vowel inventory of their description. In addition, Shen (1981) observed that although the two vowels were kept distinct among older speakers, they were in free variation among young speakers. According to Shen, the variant used most often was a vowel halfway between [ɛ] and [e] (here, transcribed as [ɛ̝]), followed by [ɛ] and then [e], with no phonological conditioning.

More quantitative data came from two language surveys carried out in the early 1980s. Shi and Jiang (1987) reported a study of 500 middle-aged speakers (age 35–55) conducted in 1983 in which 69.0% of speakers pronounced an intermediate vowel [ɛ̝] in all three lexical sets; 6.6% produced both [e] and [ɛ], mixing the lexical sets; and 24.4% manifested the previous pattern, producing [e] for MN-[ej] items only. Younger speakers showed a similar tendency to merge, only to a greater extent. Xu, Tang, and Tang's (1986) survey of 160 teenagers (age 13–14) conducted in 1980 showed that 82.5% of these young speakers had completed or were in the process of completing the merger, while only 17.5% preserved the distinction consistently.

By contrast, Svantesson (1989) found total separation of /e/ and /ɛ/ in all three speakers he examined (all male, ages mid 20s to early 50s). Svantesson's participants consistently produced MN-[ej] items with a higher vowel than MN-[an] items. However, the MN-[ej] items he used belonged to a subset of MN-[ej] items whose Mandarin counterparts contain a medial [w] before the vowel; hence, it is not clear whether the production patterns observed in these items would also hold up for other MN-[ej] items (see our Study 2). Another concern with Svantesson's study is that the speakers were living in Europe at the time of recording, which might have allowed their Shanghainese to be influenced by a different set of contact languages—both Chinese and non-Chinese—than those relevant for their peers in Shanghai.

To sum up, the most systematic studies from this body of research suggest that in the early 1980s, the merger of Shanghainese /e/ and /ɛ/ was already mainstream among middle-aged speakers and gaining momentum among younger speakers. Thus, it is reasonable to predict that by the mid to late 1980s, before the reinforcement of the Mandarin education policy, the merger would have been a pervasive trend, although the state of complete merger might not have ever been reached.

Since the early 2000s, a novel, reverse trend has been noted, wherein the two vowels seem to have become distinct again, as MN-[ej] items have readopted the [e] pronunciation (Gu 2007, Qian 2003:57). Despite disagreement about the geographic reach of the reversal, Gu (2007) and Qian (2003) agreed that the new [e] vowel showed traces of diphthongization, suggesting influence from Mandarin. However, the conclusions in these studies were based on impressionistic examination of averaged acoustic measurements, and not much detail was provided regarding the acoustic and sociolinguistic status of the change or, for that matter, its psycholinguistic basis.

In the current paper, we report the findings of a systematic examination of both the linguistic status of this sound change and the speaker-internal mechanisms that have given rise to the change. From an SLA point of view, the speakers of interest spoke Shanghainese as their L1 and acquired Mandarin as an L2. Thus, the central SLA question is: how did a feature of the L2 get transferred into the L1? Moreover, what factors have constrained this crosslinguistic influence? In order to address these questions, we first lay out a theoretical framework for the analysis of bilingual speech.

2.3. MODELING THE BILINGUAL SPEAKER. The two most influential and widely tested models of L2 speech are the Speech Learning Model (SLM; Flege 1995, 1996) and the Perceptual

Assimilation Model-L2 (PAM-L2; Best & Tyler 2007). Following from the general objectives of SLA research, both models provide insight into the acquisition of L2 phonology, including the influence of the L1 on the developing L2. However, a crucial difference between the two is that only the SLM provides a theoretical motivation for bidirectional influence between the L1 and the L2: the coexistence of L1 and L2 sounds in a shared, malleable system. According to the SLM, when a new sound is encountered by a learner, 'phonetic systems reorganize ... through the addition of new phonetic categories, or through the modification of old ones' (Flege 1995:233). That is to say, when L2 sounds are acquired, the representation of preexisting L1 sounds may change in response. One mechanism that allows for this crosslinguistic influence is EQUIVALENCE CLASSIFICATION, which causes 'similar' L1 and L2 sounds to be linked perceptually at a position-specific allophonic level and processed under the same phonetic category with merged phonetic properties. As a result of such perceptual linkage, both the L1 and L2 sounds may be produced differently than they would be by monolinguals.

A crucial condition for equivalence classification is phonological similarity. The SLM enumerates three criteria for evaluating phonological similarity between L1 and L2 sounds: identity in transcription, acoustic proximity, and perceptual similarity (Flege 1996:16–18). However, the latter two criteria are gradient and, thus, denote a construct of 'similarity' that is fluid and continuous, rather than categorical (Chang 2010). How similar is similar enough for cross-language linkage to occur (resulting in L2-to-L1 influence) is one of the questions we explore in our experiments (see §4). A further question is the linguistic level at which crosslinguistic phonological similarity is determined. For example, are the L1 and L2 compared at the level of the segment (cf. the position-specific allophonic level assumed in the SLM), the natural class, or the word?

There is abundant evidence that L2-to-L1 influence can occur between similar-sounding phones (i.e. at the level of the segment). For example, Flege (1987) investigated the production of /t/ by native speakers of American English living in Paris and native speakers of French living in Chicago, all advanced L2 learners of the ambient language. In comparison to monolingual phonetic norms, both groups' L1 speech showed L2 influence, with voice onset time (VOT) manifesting phonetic drift in the direction of the L2. The convergent nature of this L2-to-L1 influence is consistent with predictions of the SLM. As English /t/ and French /t/ are phonetically similar, they may undergo equivalence classification in a bilingual's (shared) language system, causing productions to come even closer to each other. A more recent study by Chang (2012) found evidence that such convergence occurs in novice L2 learners as well—in both VOT as well as properties of vowel production such as the fundamental frequency onset and the first formant frequency (F1).

In regard to determining crosslinguistic phonological similarity, the L2-to-L1 influence documented in Chang (2012) is noteworthy because it was found at multiple levels of phonological structure, not just at the segmental level. Thus, the L1 (English) production of the participants evinced influence from L2 (Korean) phonetic norms not only at the level of the segment (e.g. /t^h/), but also at the level of the natural class (e.g. aspirated stops) and the level of the system (e.g. global F1 over all vowels in the inventory). Furthermore, these different kinds of L2-to-L1 influence were not mutually exclusive; instead, they were often found to jointly influence participants' L1 production. These findings suggest that although crosslinguistic equivalence classification (and, thus, perceptual linkage) can occur on a segment-to-segment basis, it occurs more broadly than predicted by the position-specific allophonic focus of the SLM.

In addition to documenting various shades of L2-to-L1 influence, bilingual speech research has further shown that the amount of L2-to-L1 influence can vary within the individual as the language environment changes. Sancier and Fowler (1997) reported a study of a late L1 Portuguese-L2 English bilingual, who traveled back and forth between the US and her native Brazil on a regular basis. Acoustic analysis showed that the speaker produced Portuguese voiceless stops with longer (i.e. more English-like) VOTs after an extended stay in the US, compared with her own productions after an extended stay in Brazil. In other words, the speaker's L1 speech converged with the L2 to a greater degree after recent immersion in the L2. The authors accounted for these results in terms of three tendencies: imitation of ambient sounds, linkage between similar phones across languages, and the recency effect in memory.

In our view, Sancier and Fowler's (1997) findings may also be related to the influence of language mode. Grosjean (2001:3) defined the bilingual's language mode as 'the state of activation of the bilingual's languages and language processing mechanisms at a given point in time'. It has long been thought that bilingual speakers can switch between multiple language modes (e.g. monolingual L1, monolingual L2, bilingual) based on the communicative situation

(Clyne 1972, Grosjean 2001, Hasselmo 1970, Weinreich 1968). Bilingual speakers may, for instance, avoid using their other language and show less crosslinguistic influence when talking with monolingual speakers, but engage in more code-switching and/or manifest more crosslinguistic influence in communication with other bilinguals. Apart from the interlocutor, Grosjean listed a wide range of other factors that might influence language mode, including (but not limited to) features of the situation (e.g. physical location), the language environment (e.g. language exposed to), and the language act (e.g. communicative functions). With regard to Sancier and Fowler's findings, one may surmise that the bilingual speaker's language mode had changed after living in the US for several months. Immersion in an English language environment would have increased the activation level of English, which would have likely encouraged her to switch away from a monolingual L1 mode. This increased activation of the L2, possibly accompanied by inhibition of the L1, would have facilitated crosslinguistic influence of English on Portuguese.

2.4. RESEARCH QUESTIONS AND PREDICTIONS. As stated above, the goals of the current research were twofold: (1) to analyze the linguistic status of an ongoing sound change in Shanghainese, and (2) to explore the source of this change in bilingual processing and production. To achieve the first goal, we obtained acoustic measures of the relevant vowels produced by a sizable speaker sample spanning generations. To achieve the second goal, we adopted the general framework of the SLM, supplemented with other findings in acquisition and bilingualism.

Our main hypothesis was that the sound change originated from selective perceptual linkage between Shanghainese and Mandarin in the bilingual's language system. More specifically, we hypothesized that because of the phonetic similarity of [ɛ] and [ej] (which both contain mid front unrounded vowel nuclei), some current speakers of Shanghainese analogize Shanghainese [ɛ] to Mandarin [ej] in MN-[ej] items, causing the pronunciation of the former to drift toward the latter. On the other hand, the phonetic differences between [ɛ] and the other two Mandarin rhymes, [an] and [aj] (which both contain low vowel nuclei), are salient and relatively easily perceived; therefore, in contrast to MN-[ej] items, no drift was predicted to occur in MN-[an] or MN-[aj] items. These feature-based judgments of relative crosslinguistic similarity are consistent with published acoustic norms for the first two formants (F1 and F2) in Shanghainese and Mandarin vowels (although these norms differ in precision and reliability across studies, sometimes being based on only two speakers). Shanghainese [ɛ] is reported as having F1 values of 500–600 Hz and F2 values over 2000 Hz (for both sexes; Chen 2008, Svantesson 1989), which are close to the values reported for the vowel nucleus in Mandarin [ej] by Wu (1986): 600–700 Hz (F1) and 1800–2400 Hz (F2).[5] By contrast, formant values in Wu 1986 for the vowel nuclei in Mandarin [aj] and [an] (F1 > 800 Hz, F2 < 2000 Hz) place both of these vowels farther away from Shanghainese [ɛ].

The overall picture for Shanghainese [ɛ] is presented in Figure 1, which schematizes the bilingual lexical representations for the three lexical sets at issue. In each case, the relevant cross-language linkage (or lack thereof) is posited to occur at the lexical-phonetic level. Note that cross-language linkage at the lexical-phonetic level does not follow from the SLM, which neither excludes nor predicts this possibility. Rather, it follows from the occurrence of cross-language linkage beyond the allophonic level (Chang 2012) as well as holistic lexical representations (Caramazza et al. 1988, Ferguson & Farwell 1975, Metsala & Walley 1998). Given that crosslinguistic comparison occurs at higher linguistic levels and one level of linguistic representation appears to be a 'whole word' representation, it stands to reason that cross-language linkage may also occur between corresponding word forms. This type of cross-language linkage in one lexical set (MN-[ej]), but not the other two, creates the potential for reestablishing a phonemic distinction between two mid front vowels.

<INSERT FIGURE 1>

The hypothesis of perceptual linkage between the Shanghainese and Mandarin word forms in the MN-[ej] lexical set generates several predictions that distinguish it from a possible alternative hypothesis attributing the sound change to the residue of an incomplete merger (i.e. surviving [e] pronunciations). First, if the ongoing change is indeed due to influence from Mandarin, the new vowel in MN-[ej] items should bear traces of diphthongization—a clear feature of the Mandarin [ej] vowel—as noted in previous studies. Second, in accordance with the SLM, phonological similarity should play a critical role in conditioning cross-language convergence. Since the hypothesized cross-language linkage is established at the lexical-phonetic level, the presence and strength of the linkage ought to be influenced by similarity in the whole syllable, including onset, medial, and rhyme. Third, following from both the SLM and the language mode theory, the sound change should be more evident among speakers born after 1980, who have more experience with Mandarin from an early age, than among older generations. Younger speakers not only have more authentic representations of Mandarin vowels than older speakers, but are also more subject to Mandarin-to-Shanghainese influence because Mandarin is presumably more activated in their linguistic systems. Finally, along the same lines, the sound change should be more evident when a speaker is operating in a bilingual mode (in which Mandarin is highly activated) than when a speaker is operating in a monolingual Shanghainese mode (in which Mandarin is maximally deactivated).

In order to test these predictions, we conducted two production experiments with two generations of bilingual participants: a sentence-reading experiment and an auditory translation experiment. The two experiments were designed to mimic the monolingual Shanghainese mode and the bilingual mode, respectively. To examine crosslinguistic similarity as a gradient variable influencing phonetic drift, we tested three types of MN-[ej] items differing in degree of crosslinguistic similarity and analyzed them in separate studies. Study 1 examined MN-[ej] items that were identical to their Mandarin counterparts with respect to onset consonant and maximally similar with respect to syllable structure (i.e. onset + rhyme)—the highest possible level of similarity for our purposes. Study 2 focused on MN-[ej] items that were less similar to their Mandarin counterpart with respect to syllable structure. Study 3 examined MN-[ej] items that were less similar to their Mandarin counterpart with respect to onset.

Before we move on to describing the experimental procedure and results of Studies 1–3, it should be noted that while we hypothesize the Shanghainese sound change to be contact-induced, we make no claim regarding its subsequent spread throughout the lexicon and in the speaker population, which may not depend on language contact. Thus, our hypothesis does not preclude Shanghainese words without Mandarin cognates or Shanghainese speakers without knowledge of Mandarin from showing the sound change, although it is reasonable to predict that such words and such speakers will be less likely to undergo the change and, insofar as they do, will show the change later and probably to a lesser degree (in comparison to words with cognates in Mandarin and to bilingual speakers, respectively). Testing this prediction is beyond the scope of the current paper, as it requires different experimental methods than those used in the current study (see §3.2 for further discussion).

3. Methods.

3.1. Participants.

3.1. PARTICIPANTS. A total of 24 native speakers of Shanghainese—all born, raised, and resident most of their life in Shanghai—participated in the production experiments. The participants were recruited in the form of 12 parent-child pairs to control for immigration history and linguistic variation across urban districts. The older participants (i.e. the parents) were born between the late 1940s and mid 1950s (7 female; mean age = 59.2 yr, s.d. = 3.2) and were thus roughly peers with the middle-aged speakers in Xu & Tang 1988. The younger participants (i.e. the children) were born between the late 1970s and mid 1980s (8 female; mean age = 29.8 yr, s.d. = 3.9) and went through most of their schooling with a strict Mandarin-only policy. All participants reported speaking Mandarin as an additional language. None had a history of speech or hearing disorders.

A separate group of 23 Shanghainese speakers (15 female; mean age = 27.5 yr, s.d. = 6.9) participated in an online word-frequency rating task meant to help assign experimental items to frequency conditions. All the raters reported being born in Shanghai, as well as a long residential history in Shanghai (mean length = 19.1 yr, s.d. = 7.8), and passed a language screening test by correctly translating a series of Shanghainese auditory forms into English.

3.2. Materials.

TEST MATERIALS. A total of 27 monosyllabic, monomorphemic lexical items from three lexical sets (MN-[aj], MN-[an], MN-[ej]) were examined in the production experiments; 18 of these were analyzed in Study 1, 9 in Study 2, and 12 in Study 3 (with some items shared among the studies). Since most of these items cannot be used alone, each was embedded in a multisyllabic (generally bisyllabic, only one quadrasyllabic) compound word in the final position. An example compound is shown in 1, with the critical item bolded. The embedding compounds were classified into two frequency bands (High, Low) based on subjective ratings pooled from the online rating task. On a scale of 0–4 (0 = unknown, 1 = very infrequent, 2 = infrequent, 3 = frequent, 4 = highly frequent), high-frequency items received an average rating of 3.3 (s.d. = 0.5), while low-frequency items received an average rating of 1.9 (s.d. = 0.5). The difference between the two frequency bands was significant ($t(27) = 7.29$, $p < .001$).

(1) 外 **滩**
ŋɑ˧ **tʰɛ˥**
outside **waterfront**
'the Bund' (name of a waterfront area in central Shanghai)

Critical items (and their embedding compounds) were carefully chosen to control for phonetic context and usage frequency. The use of high/mid front vowels in syllables preceding the critical items was avoided in order to prevent vowel-to-vowel assimilation that might cause raising in the critical vowel. In addition, onset consonants were maximally matched across lexical sets to control for consonant-to-vowel coarticulatory effects. Due to lexical gaps, all critical items had a bilabial or coronal onset that belonged to one of four categories: lateral approximant /l/ (L); voiceless stops with long-lag VOT /p^h t^h/ (PHTH); voiced and voiceless stops with short-lag VOT /b p t/ (PTB); and sibilant fricative /s/ (S). Stops were grouped by VOT lag because (1) gaps in the syllabary made it impossible to use the exact same plosives for all of the relevant lexical sets, and (2) in the given mid front vowel context, differences in VOT lag have a greater effect on onset vowel formants than differences in place of articulation. Items with the voiced onset /b/ were only used in Study 3.

The selection of critical items was further constrained by the nature of the experimental tasks. For example, sentence reading made it difficult to reliably elicit native Shanghainese words, which generally lack standard and widely-accepted orthographic forms; consequently, we only included items that had cognates in Mandarin. Thus, the current study is concerned only with the part of the Shanghainese lexicon that has Mandarin cognates, which is exactly the part of the lexicon we hypothesize to be the birthplace of the Shanghainese merger reversal.

Apart from the critical items discussed above, the test materials also included 32 filler compounds of similar lengths that had Mandarin cognates but did not contain the critical vowels (Shanghainese [e] and [ɛ]; Mandarin [aj], [an], and [ej]) in the pronunciation of either language, as well as 3 critical compounds that were dropped from final analysis because of a repeated critical character (*n* = 1), a frequency mismatch (*n* = 1), or a high rate (> 50%) of mispronunciation among the participants (*n* = 1). Thus, the test materials consisted of 62 (27 + 32 + 3) compounds in total.

EXPERIMENTAL STIMULI. The stimuli in the sentence-reading experiment comprised 62 sentences, one for each test compound. In a critical sentence stimulus, the critical compound (hence, also the critical vowel) always appeared in prepausal (i.e. clause-final) or near prepausal position (followed only by the sentence-end particle *le*) to control for prosodic effects on vowel production. Sentences containing a critical item (e.g. 2, where the critical compound is bolded) typically had 2–3 clauses separated by commas and were 10–26 syllables long (mean = 18.1 syllables, s.d. = 4.1). Sentences for the filler items were created in the same fashion with similar lengths (mean = 19.0 syllables, s.d. = 3.6), which did not differ significantly from the lengths of the sentences containing critical items ($t(52) = -0.93$, $p = .36$).

(2)	我	以为	伊	去了	南京路，	原来	是	去了	**外滩。**
	ŋu˧	ʔi˥ we˥	ʔi˧	tʰɕi˦ lə˥	nø˧ tɕin˥ lu˧	ɲy˧ lɛ˥	z̩˧	tʰɕi˧ lə˥	**ŋɑ˧ tʰɛ˥**
	I	think	s/he	go.ASPECT	Nanjing Rd.	actually	be	go.ASPECT	**the Bund**

‘I thought that s/he went to Nanjing Road. It turned out that (s/he) went to the Bund.’

Given that Shanghainese is used almost exclusively as a spoken language now, one might have concerns that the sentence-reading experiment could elicit unnatural Shanghainese pronunciations (or even Mandarin pronunciations) because participants might find it difficult to read written Chinese aloud in Shanghainese. To address these concerns, we employed several measures to facilitate the production of natural, fluent Shanghainese in the reading task. First, the task was self-paced to allow participants enough time to prepare for each utterance. Second, the structures of the sentences were relatively simple and also similar to the structures of translation equivalents in Mandarin, while linguistic features used only in Mandarin and not in Shanghainese were avoided. Finally, native Shanghainese words with conventionalized and/or relatively straightforward orthography (often based on characters that are virtually homophonic in Shanghainese and Mandarin) were used whenever possible in order to make the sentences sound more natural and authentic.

To check that the reading task was successful in eliciting natural Shanghainese, we also conducted a post-hoc rating study with two trained linguists, native speakers of Shanghainese who were peers in age, respectively, with the younger and older groups in the main experiments. Naïve to the true purpose of the rating study, both linguists listened to a random subset of the sentence recordings (two from each speaker) and rated each recording on the naturalness of the speech and the authenticity of the pronunciation using a 5-point scale (1 = very unnatural/accented, 5 = very natural/authentic). Their average ratings were high: 4.12 (s.d. = 1.01) for naturalness and 4.17 (s.d. = 1.15) for authenticity. Both raters reported hearing only Shanghainese, and no Mandarin productions, in the recordings, suggesting that the sentence-reading experiment did indeed elicit naturally produced, authentic Shanghainese.

As for the Mandarin-to-Shanghainese translation experiment, the stimuli in this experiment comprised 62 audio recordings of the Mandarin counterparts of the test compounds (e.g. Mandarin [waj˥˩ t^han˥] for 外滩 'the Bund'). The recordings were of a female native speaker of Beijing Mandarin in her 30s and were made at 48 kHz (stereo) and 16 bps in a sound-attenuated room with a Marantz PMD660 recorder and its internal microphone.

3.3. PROCEDURE.

FREQUENCY RATING TASK. Because suitable Shanghainese corpora and frequency dictionaries were not available, we conducted an online frequency rating task in order to gather frequency data on potential items for the production experiments. In this task, raters heard 100 prerecorded Shanghainese compounds (68 potential critical items and 32 fillers) read by a female native speaker and were instructed to estimate the usage frequency of each compound on a scale of 0–4. An English translation for each item was provided on screen for disambiguation; however, Chinese orthography was not presented so as to prevent interference from Mandarin frequencies. Results from the rating task were used to select critical items for the production experiments. All else being equal, compounds with the most extreme (i.e. highest or lowest) frequency ratings were chosen in order to maximally differentiate the two frequency conditions.

PRODUCTION EXPERIMENTS. The production experiments took place in a quiet, closed room at the home of either the participant or the experimenter (i.e. the first author). All conversations between the participant and the experimenter, including the provision of instructions, were conducted in Shanghainese. The sequence of experiments was sentence reading then translation, with a break of about 20 minutes in between for a language background survey. Both experiments were presented in DMDX (Forster 2008) and entirely self-paced. During the experiments, the participant sat at a desk, facing a Lenovo ThinkPad (T400) laptop and wearing an AKG C420 head-mounted condenser microphone connected to a Marantz PMD660 recorder. The experimenter sat beside the participant throughout the experiments to provide necessary technical assistance, since quite a few older participants were unfamiliar with the computerized setting. A complete session typically lasted about one hour.

On each trial of the reading experiment, the participant was presented with a sentence on screen to read aloud in Shanghainese. On each trial of the translation experiment, the participant was presented with an auditory Mandarin stimulus through computer speakers to translate out loud into Shanghainese within the carrier sentence [gəʔ˧ {tɕiɔ˦/z̩˧} ...] 'This is (called) ... '. The written form of the stimulus was also shown on screen 250 milliseconds after the audio ended in case the participant had difficulty identifying what s/he had heard; however, participants were instructed not to look at the screen unless they needed to (which rarely occurred in the experimental sessions). Thus, participants' oral responses were primarily cued by the auditory stimuli, with minimal influence from orthography.

Both experiments began with three practice trials (not included in the test stimuli) and continued with three test blocks, each of which iterated over all 62 test stimuli in a randomized order, resulting in three presentations of each stimulus. Post-experiment surveys showed that most participants noticed that the compounds for translation were contained in the sentences for reading, but none paid special attention to the critical vowels of interest.

3.4. ANALYSIS. Acoustic analysis of the recordings was done in Praat (Boersma & Weenink 2011). The onset and offset of the critical vowels were marked by hand, and then F1 and F2 were measured using linear predictive coding analysis at the 20% (start) and 80% (end) points of the vowel. Spectrograms of all tokens were inspected visually for accuracy of formant tracking, and errors were corrected manually by examining spectral slices at the appropriate time points. Formant values were converted to the mel scale, a perceptual scale of pitch (Stevens et al. 1937), to obtain a perceptual picture of formant trajectories. For each vowel token, there were thus four formant measures (F1Start, F2Start, F1End, F2End). In addition, a binary measure indicating diphthongization toward [i] (hereafter, Diphthong) was derived on the basis of the formant measures. If a token moved up and front (i.e. F1End < F1Start, F2End > F2Start), Diphthong was coded as 1, else as 0.

The five phonetic measures (F1Start, F2Start, F1End, F2End, Diphthong) were examined as outcome variables in separate mixed-effects regression models for each task (reading, translation) in each study: linear mixed (LM) models on the formant measures and generalized linear mixed (GLM) models on Diphthong. Each regression model initially contained two random effects—by-item and by-family intercepts (which controlled for random variation among test items and speaker families)—and a set of fixed-effect terms involving an item's lexical set (LexSet), type of onset consonant (Onset), embedding compound frequency (Frq), speaker age (Age), speaker sex (Sex), and test block (Block). The critical predictor terms were LexSet and its interaction with Age (LexSet×Age), reflecting age-related sound change patterns (if any) in the data. The control predictors were LexSet×Sex, LexSet×Frq, Block, and Onset; the first two controlled for variation due to speaker sex and usage frequency that may interact with LexSet, while the last two controlled for effects of repetition and the preceding consonant on vowel production. Initial model structure is shown in Table 2.

<INSERT TABLE 2>

The fixed-effect predictors were coded as follows. Both LexSet and Age were categorical variables with treatment contrast coding, and the reference levels were set to levels that facilitated model interpretation (see §4 for more details).[6] As for the control predictors, Block was an ordered categorical variable with orthogonal polynomial coding and had two components (Block.L, Block.Q) representing the linear and quadratic trends of Block, respectively. The other control factors (Frq, Onset, and Sex) were all categorical variables with sum contrast coding, where each level was compared to the grand mean.[7] Table 3 summarizes the properties of the fixed-effect predictors in each study.

<INSERT TABLE 3>

The data from the different studies (Studies 1, 2, 3) were modeled separately due to the unbalanced nature of the composite dataset (see Table 3). Furthermore, data from different tasks (sentence reading, translation) in the same study were also modeled separately, in order to avoid terms of high-order interaction (e.g. LexSet×Age×Experiment), which are difficult to interpret. Task-related differences were examined instead by comparing across models. To prevent interference from obviously nonsignificant predictors, the fixed-effect predictors underwent two rounds of backward elimination in each model. In the first round, only nonsignificant interaction terms were removed, resulting in an updated, simplified model. In the second round, all remaining main and interaction terms were subject to elimination. In each elimination step, if the change in log likelihood of the model due to the exclusion of a certain predictor term was clearly nonsignificant ($p(\chi^2) > .05$), the term was removed from the final model.[8] All statistical analyses were carried out using the lme4 package (Bates et al. 2011) in R (R Core Team 2012).

4. RESULTS.

4.1. STUDY 1: CONFIRMATION OF CHANGE. The test items in Study 1 were 18 monosyllabic CV items: 6 triplets contrasting three levels of LexSet (MN-[aj], MN-[an], MN-[ej]). All test items were identical to their Mandarin counterparts in onset consonant and maximally similar in syllable structure.[9] Three levels of Onset (L, PHTH, PTB with only /p t/) and two levels of Frq (High, Low) were balanced among the six triplets, resulting in a 3×3×2 design. All the test items in Study 1 (as well as Studies 2–3) are listed in the online supplementary materials. With three levels of Block (1, 2, 3), two levels of Age (Young, Old), and 12 speakers in each age group, the dataset for each experiment included about 1296 tokens (18 items×3 blocks×2 ages×12 speakers), with occasional missing tokens (< 4%) due to unrecognized words or corrupted sound files.

A total of 10 mixed models were constructed, one for each of the five dependent variables (F1Start, F2Start, F1End, F2End, Diphthong) in each of the two experiments (sentence reading, translation), following the steps described in §3.4. The reference levels of LexSet and Age were set to MN-[aj] and Young, respectively. Thus, a main effect of LexSet when LexSet=MN-[ej] (corresponding to the coefficient $\beta_{\text{MN-[ej]}}$) represents the predicted difference between MN-[ej]

items and MN-[aj] items in YOUNGER speakers when all other factors are controlled, and the summation of this main effect and the LexSet×Age interaction effect when LexSet = MN-[ej] and Age = Old (i.e. $\beta_{\text{MN-[ej]}} + \beta_{\text{MN-[ej]:Old}}$) represents the predicted difference between MN-[ej] items and MN-[aj] items in OLDER speakers. If $\beta_{\text{MN-[ej]}}$ is significantly different from zero but $\beta_{\text{MN-[ej]:Old}}$ is not, this indicates that the difference between MN-[ej] and MN-[aj] items exists in both age groups to a similar degree. The significance of $\beta_{\text{MN-[ej]}} + \beta_{\text{MN-[ej]:Old}}$ is not directly examined by the models; thus, when the significance of the summed value is in doubt (e.g. when $\beta_{\text{MN-[ej]}}$ is not significantly different from zero but $\beta_{\text{MN-[ej]:Old}}$ is, or when $\beta_{\text{MN-[ej]}}$ and $\beta_{\text{MN-[ej]:Old}}$ are similar in magnitude but have opposite signs), results from alternative models where Old is set as the reference level of Age are examined to determine the significance of the MN-[ej] vs. MN-[aj] difference in older speakers.

Significance of the predictor terms in the LM models on formant measures was determined by p_{MCMC} values, calculated based on the posterior distribution of model parameters generated by the Markov Chain Monte Carlo (MCMC) sampling procedure (10,000 samples; see Baayen et al. 2008 for a description of the procedure). Predictor terms with p_{MCMC} values smaller than .01 were considered to be statistically significant and p_{MCMC} values between .01 and .05 marginally significant, while p_{MCMC} values greater than .05 were considered nonsignificant (n.s.). The critical statistics in each model are cited in the text below. A full report of model parameters for fixed-effect terms for Study 1 (as well as Studies 2–3) is provided in the online supplementary materials.

READING EXPERIMENT. Model results for the reading experiment showed no effect of LexSet or any interaction involving LexSet on formant values near vowel onset (i.e. F1Start, F2Start), suggesting that formant onsets were similar across lexical sets. Near the offset of the vowel, however, there were significant effects of LexSet on both formant values (i.e. F1End, F2End) as well as a significant LexSet×Age interaction for F2End. In particular, all else being equal, MN-[ej] items ended with lower F1End ($\beta_{\text{MN-[ej]}} = -74.17$, $t = -4.87$, $p_{MCMC} < .001$; no LexSet×Age interaction) and higher F2End ($\beta_{\text{MN-[ej]}} = 66.74$, $t = 4.20$, $p_{MCMC} < .001$; $\beta_{\text{MN-[ej]:Old}} = -24.95$, $t = -3.21$, $p_{MCMC} = .002$) than MN-[aj] items in both age groups, and the increase in F2End was significantly smaller in older speakers than in younger speakers, as shown by the sign of $\beta_{\text{MN-[ej]:Old}}$. No reliable difference was detected between MN-[an] and MN-[aj] items (F1End: $\beta_{\text{MN-[an]}} = -5.65$, $t = -0.37$, n.s.; no LexSet×Age interaction. F2End: $\beta_{\text{MN-[an]}} = 21.87$, $t = 1.38$, p_{MCMC} = n.s.; $\beta_{\text{MN-[an]:Old}} = -18.20$, $t = -2.32$, $p_{MCMC} = .020$).[10]

Figure 2a plots the grand means (averaged over individual speakers' means) of F1Start and F1End and of F2Start and F2End in the form of vectors in the F1-F2 plane. The data are separated by LexSet and Age; since the two age groups had a different sex composition and biological sex strongly affects vowel formants, the data are further separated by Sex. On average, female speakers' MN-[ej] vowels were about 72–75 mel lower in F1End and 37–52 mel higher in F2End than their vowels in the other lexical sets, while male speakers' MN-[ej] vowels were about 58–77 mel lower in F1End and 44–55 mel higher in F2End than their vowels in the other lexical sets.

<INSERT FIGURE 2>

Thus, compared with the other two lexical sets, MN-[ej] items tended to start in roughly the same position but end in a higher and more front position, suggesting that the vowel in MN-[ej] items tended to be diphthongized toward [i]. Results from the GLM model confirmed that MN-[ej] items were significantly more likely to be diphthongized toward [i] than the baseline MN-[aj] items across age groups ($\beta_{\text{MN-[ej]}} = 2.42$, $z = 9.75$, $p(|z|) < .001$; no LexSet×Age interaction). The model also suggested that MN-[an] items had a slightly higher rate of diphthongization than MN-[aj] items ($\beta_{\text{MN-[an]}} = 0.50$, $z = 2.00$, $p(|z|) = .045$), but this effect was small and marginally significant. Overall, MN-[ej] items were produced with diphthongized vowels 59.2% of the time, whereas MN-[aj] and MN-[an] items were each diphthongized less than 30% of the time (Figure 3a).

<INSERT FIGURE 3>

TRANSLATION EXPERIMENT. Model results for the translation experiment showed trends similar to those found in the reading experiment, only more pronounced. As in the reading experiment, neither LexSet nor its interaction with any other predictor had an effect on F1Start or F2Start; however, both LexSet and LexSet×Age significantly conditioned F1End and F2End. Younger speakers' vowels in MN-[ej] items had lower F1End ($\beta_{\text{MN-[ej]}} = -130.14$, $t = -6.60$, $p_{MCMC} < .001$) and higher F2End ($\beta_{\text{MN-[ej]}} = 86.36$, $t = 7.79$, $p_{MCMC} = .001$) than their vowels in

MN-[aj] items. Similar effects were found with older speakers, but reduced in size (F1End: $\beta_{\text{MN-[ej]:Old}} = 34.99$, $t = 3.14$, $p_{MCMC} = .002$; F2End: $\beta_{\text{MN-[ej]:Old}} = -43.24$, $t = -5.50$, $p_{MCMC} < .001$). No significant difference was found between MN-[an] items and MN-[aj] items in either formant measure in either age group (all $|t|$'s < 1, all p_{MCMC}'s ≥ .4).

Comparison of the two experiments showed that cross-lexical set differences in F1End and F2End were greater in translation than in reading. As shown in Figure 2b (cf. Figure 2a), female speakers' MN-[ej] production in translation was on average 117–119 mel lower in F1End (cf. 72–75 in reading) and 61–62 mel higher in F2End (cf. 37–52 in reading) than their production of the other lexical sets, while male speakers' MN-[ej] production was 105 mel lower in F1End (cf. 58–77 in reading) and 67–68 mel higher in F2End (cf. 44–55 in reading) than their production of the other lexical sets. The LexSet×Age interaction effects were also more pronounced in translation than in reading. As mentioned above, only one formant measure (F2End) showed a significant LexSet×Age interaction in reading, and the size of this effect was larger in translation ($\beta_{\text{MN-[ej]:Old}} = -24.95$ in reading; $\beta_{\text{MN-[ej]:Old}} = -43.24$ in translation).

As in the reading experiment, MN-[ej] items showed a strong trend of diphthongization toward [i] in the translation experiment. Vowels in MN-[ej] items were significantly more likely to move up and front in the F1-F2 plane compared with vowels in MN-[aj] items, and more so in younger speakers ($\beta_{\text{MN-[ej]}} = 3.12$, $z = 11.60$, $p(|z|) < .001$) than in older speakers ($\beta_{\text{MN-[ej]:Old}} = -0.91$, $z = -2.64$, $p(|z|) = .008$). No difference, however, was detected between MN-[an] and MN-[aj] items (all $|z|$'s < 1.6, all $p(|z|)$'s > .1). On average, MN-[ej] items were diphthongized almost 80% of the time by younger speakers and 70% of the time by older speakers, whereas MN-[aj] and MN-[an] items were diphthongized less than 30% of the time for both age groups (Figure 3b).

To sum up, results from the translation experiment replicated the major findings of the reading experiment. Compared with the other lexical sets, MN-[ej] items tended to end higher and more front in the F1-F2 plane and showed a significant trend of diphthongization toward [i]. These patterns were more pronounced in younger than in older speakers, and more pronounced in translation than in reading.

With regard to the differences in effect size found between experiments, it is important to consider differences in aspects of the experimental task. The reading experiment elicited productions of relatively long sentences with no preassigned focus, whereas the translation experiment elicited productions of a short carrier sentence, which clearly put focus on the target compound. Unsurprisingly, post-hoc analysis showed vowel tokens in translation were longer than those in reading (see Table 4; paired t-test: $t(427) = 16.44$, $p < .001$). Given the well-established correlation between duration and vowel undershoot/overshoot (Lindblom 1990, Moon & Lindblom 1994), it is possible that vowel productions in translation extended across a larger vowel space than those in reading, augmenting the distance between different vowels and, therefore, the distinction between MN-[ej] items and the other lexical sets in translation.

<INSERT TABLE 4>

Nevertheless, closer scrutiny of the data suggests that vowel duration differences cannot fully account for the observed patterns of cross-experiment variation, because the patterns are not completely compatible with the expected effect of duration. Importantly, younger speakers' vowels were significantly shorter than older speakers' in translation (all p's < .05); however, despite the shorter durations, their MN-[ej] vowels (but not the vowels of the other lexical sets) had more extreme ending positions and were diphthongized toward [i] much more often than older speakers' (see Figure 2b). To put it a different way, there was an Experiment×LexSet×Age interaction that could not be explained by generally longer vowel durations in translation, since it was actually the shorter vowels—those MN-[ej] vowels produced by faster-speaking younger speakers—that showed the most extreme acoustic excursions.

How, then, can we account for the cross-experiment variation that was conditioned by LexSet and Age? We argue that this variation stems from the difference in language mode between experiments. Whereas the reading experiment elicited a mainly monolingual Shanghainese mode with minimal activation of Mandarin, the translation experiment required the participant to operate in a bilingual mode by using a cross-language auditory priming paradigm. Recall that we predicted that if the vowel change in MN-[ej] items was due to contact with Mandarin, it would be more evident in bilingual mode than in monolingual Shanghainese mode; we predicted further that a bilingual mode would result in more evidence of the vowel change for younger speakers (who were more balanced bilinguals than older speakers, with a generally higher activation level of Mandarin). The Experiment×LexSet×Age interaction in the observed direction is, therefore, consistent with our initial predictions.

CONTROL FACTORS. In addition to the critical predictors, a number of control factors were found to have a significant main effect on vowel formants, mostly in the expected direction. For example, female speakers consistently produced higher-frequency formants than male speakers. Many of the models also suggested that vowels following /l/ tended to have higher F1 but lower F2 than vowels following voiceless plosives (/p^h t^h p t/). Speaker age, however, had an unexpected effect on the phonetic measures. Although not hypothesized to globally influence vowel formants, age was found to have an effect in many of the models, especially in the reading experiment. Older speakers' vowels in reading tended overall to be lower in F1 (F1Start: $\beta_{\text{Old}} = -54.20$, $t = -15.55$, $p_{MCMC} < .001$; no LexSet×Age interaction. F1End: $\beta_{\text{Old}} = -31.73$, $t = -7.59$, $p_{MCMC} < .001$; no LexSet×Age interaction) and higher in F2 (F2Start: $\beta_{\text{Old}} = 26.13$, $t = 8.52$, $p_{MCMC} < .001$; no LexSet×Age interaction. F2End: $\beta_{\text{Old}} = 29.09$, $t = 5.21$, $p_{MCMC} < .001$; $\beta_{\text{MN-[ej]:Old}} = -24.95$, $t = -3.21$, $p_{MCMC} = .002$; $\beta_{\text{MN-[an]:Old}} = -18.20$, $t = -2.32$, $p_{MCMC} = .019$) than those of younger speakers. That is to say, older speakers' production of the target vowels was more peripheral than younger speakers'. Since vowel duration did not differ between the two age groups in reading (all *p*'s > .05), the most likely explanation is that, compared to younger speakers, older speakers were able to implement more extreme vowel articulations in relatively fast, connected speech due to their more extensive prior articulatory experience with these Shanghainese vowels.

SUMMARY. To conclude, Study 1 provided evidence that the vowel in MN-[ej] items is becoming distinct from the vowels in MN-[aj] and MN-[an] items by having a higher and more fronted ending position in the F1-F2 plane and a greater chance of being diphthongized toward [i]. The sound change was more advanced in younger than in older speakers, and in bilingual than in monolingual Shanghainese mode. These findings are consistent with our hypothesis that the source of this sound change is crosslinguistic influence from Mandarin arising from a perceptual linkage between word forms containing Shanghainese [ɛ] and Mandarin [ej]; in contrast, they do not follow from the alternative hypothesis attributing the sound change to remnants of [e] left from an incomplete merger. In Study 2 and Study 3, we examined the dynamics of this sound change in more detail by investigating how degrees of crosslinguistic phonological (dis)similarity would condition the observed crosslinguistic influence.

4.2. STUDY 2: DISSIMILARITY IN SYLLABLE STRUCTURE. The goal of Study 2 was to explore how structural similarity of the whole syllable conditions the sound change under investigation. In particular, we were interested in whether Shanghainese words that are less similar to their Mandarin counterparts in syllable structure have undergone the sound change to the same degree as those that are maximally similar to their Mandarin counterparts in syllable structure. To this end, Study 2 focused on a subset of MN-[ej] items that have a CV (i.e. /Cɛ/) structure, but correspond to Mandarin counterparts that have a CGVG (i.e. /Cwej/) structure, with a medial approximant [w] before the vowel. Consequently, this lexical set is characterized by a lower degree of crosslinguistic phonological similarity than the regular MN-[ej] items examined in Study 1. For clarity, we refer to this subset of MN-[ej] items as structure-mismatched MN-[ej] items (‘structure-mismatched items’ for short), as distinct from the regular (i.e. maximally structure-matched) MN-[ej] items in Study 1.[11]

Due to the existence of lexical gaps and the scarcity of appropriate test items, the word list in Study 2 had a small number of items and was not completely balanced. The critical test items were four structure-mismatched items, which were all embedded in high-frequency compounds and ranged over three levels of Onset (PHTH with only /t^h/, PTB with only /t/, S). These items were compared to two regular MN-[ej] items and three MN-[aj] items with matching onsets and frequencies. With three levels of Block (1, 2, 3), two levels of Age (Young, Old), and 12 speakers in each age group, the dataset for each experiment in Study 2 included about 648 tokens (9 items×3 blocks×2 ages×12 speakers), with approximately 3% missing due to unrecognized words or corrupted sound files.

As in Study 1, a set of 10 LM and GLM models was constructed to compare the vowel production of the three lexical sets. The modeling procedure and coding schemes for predictor variables were the same as in Study 1, except for two changes. First, Frq was not included as a predictor because all test items in Study 2 were embedded in high-frequency compounds; second, structure-mismatched items were set as the reference level of LexSet for easy comparison with both regular MN-[ej] and MN-[aj] items (see Tables 2–3).

READING VS. TRANSLATION. The final models showed no consistent overall or age-related differences between structure-mismatched items and the other lexical sets in terms of F1Start or F2Start in either experiment; however, more cross-lexical set variation appeared toward the offset of the vowel in both experiments (Figure 4). In reading, there was some evidence that the ending position of vowels in structure-mismatched items was somewhere between that of MN-[aj] items and that of regular MN-[ej] items. This was most evident in younger speakers' F1End, which was higher in MN-[aj] items ($\beta_{\text{MN-[aj]}} = 78.16$, $t = 7.42$, $p_{MCMC} < .001$) but lower in regular MN-[ej] items ($\beta_{\text{MN-[ej]\_regular}} = -31.82$, $t = -2.65$, $p_{MCMC} = .036$) than in structure-mismatched items. By contrast, cross-lexical set differences in F1End were significantly reduced or even eliminated among older speakers ($\beta_{\text{MN-[aj]:Old}} = -34.64$, $t = -2.46$, $p_{MCMC} = .015$; $\beta_{\text{MN-[ej]:Old}} = 37.65$, $t = 2.39$, $p_{MCMC} = .016$).[12] As for F2End in reading, there was a slight trend for younger speakers to have lower F2End in MN-[aj] items than in structure-mismatched items ($\beta_{\text{MN-[aj]}} = -52.35$, $t = -2.02$, $p_{MCMC} = .042$), but the difference was reduced among older speakers ($\beta_{\text{MN-[aj]:Old}} = 38.70$, $t = 3.87$, $p_{MCMC} < .001$). No F2End difference was detected between regular and structure-mismatched MN-[ej] items in either age group (all $|t|$'s ≤ 1.5, all p_{MCMC}'s $> .1$).

<INSERT FIGURE 4>

In contrast to reading, in translation both age groups' MN-[aj] items ended with significantly higher F1 ($\beta_{\text{MN-[aj]}} = 112.90$, $t = 10.55$, $p_{MCMC} < .001$; $\beta_{\text{MN-[aj]:Old}} = -41.66$, $t = -2.77$, $p_{MCMC} = .007$) and lower F2 ($\beta_{\text{MN-[aj]}} = -87.78$, $t = -9.60$, $p_{MCMC} < .001$; $\beta_{\text{MN-[aj]:Old}} = 47.67$, $t = 5.20$, $p_{MCMC} < .001$) than their structure-mismatched items, while no differences were found between regular and structure-mismatched MN-[ej] items (all $|t|$'s < 2, all p_{MCMC}'s $> .5$). Taken together, the results of both experiments indicated that just like regular MN-[ej] items, structure-mismatched items tended to end in a higher (i.e. lower F1End) and more front (i.e. higher F2End) position in the F1-F2 plane than MN-[aj] items.

Along the same lines, structure-mismatched MN-[ej] items were more likely to be diphthongized toward [i] than MN-[aj] items for both age groups in both experiments (Reading: $\beta_{\text{MN-[aj]}} = -2.53$, $z = -5.69$, $p(|z|) < .001$; $\beta_{\text{MN-[aj]:Old}} = 1.46$, $z = 2.57$, $p(|z|) = .010$. Translation: $\beta_{\text{MN-[aj]}} = -3.34$, $z = -8.76$, $p(|z|) < .001$; $\beta_{\text{MN-[aj]:Old}} = 1.36$, $z = 2.73$, $p(|z|) = .006$); however, regular MN-[ej] items were even more likely to be diphthongized, at least in reading (Reading:

$\beta_{\text{MN-[ej]\_regular}}$ = 1.10, z = 2.92, $p(|z|)$ = .004; $\beta_{\text{MN-[ej]\_regular:Old}}$ = -0.71, z = -1.49, n.s. Translation: $\beta_{\text{MN-[ej]\_regular}}$ = 0.50, z = 1.33, n.s.; $\beta_{\text{MN-[ej]\_regular:Old}}$ = -0.11, z = -0.23, n.s.). The average rates of diphthongization in structure-mismatched items (Reading: 44% - Young, 26% - Old; Translation: 70% - Young, 52% - Old) were also higher than those of MN-[aj] items (Reading: 8% - Young, 12% - Old; Translation: 12% - Young, 17% - Old) but lower than those of regular MN-[ej] items (Reading: 72% - Young, 39% - Old; Translation: 78% - Young, 60% - Old). In other words, structure-mismatched items tended to be diphthongized at rates intermediate between MN-[aj] items (which, along with MN-[an] items, showed the lowest rates of diphthongization) and regular MN-[ej] items (which showed the highest rates of diphthongization). This variation is shown in Figure 5.

<INSERT FIGURE 5>

CONTROL FACTORS. Just as in Study 1, models in Study 2 revealed some effects of the control factors. Female speakers consistently produced higher-frequency formants than male speakers; older speakers overall produced more peripheral vowels (lower F1 and higher F2) than younger speakers; and items with aspirated plosive onsets (/p^h t^h/) had higher F1 than items with unaspirated plosive onsets (/p t/) or /s/ onset.

SUMMARY. To sum up, the results of Study 2 suggested that structure-mismatched items have undergone the same type of vowel drift as regular MN-[ej] items, but not to the same degree. Compared to regular MN-[ej] items, structure-mismatched items tended to show less drift, both with respect to formant measures and rates of diphthongization. In particular, structure-mismatched items showed a statistically significant difference from both regular MN-[ej] and MN-[aj] items on two measures: (1) the gradient measure of F1End (in reading), and (2) the categorical measure of Diphthong (in reading). These findings accord with the difference in crosslinguistic phonological similarity between the two lexical sets: structure-mismatched items are relatively less similar to their Mandarin counterparts than are regular MN-[ej] items, and correspondingly they show less drift toward Mandarin, lagging behind in the sound change.

On the whole, however, structure-mismatched items patterned closely with regular MN-[ej] items (not differing statistically on eight out of ten measures), so one may wonder whether the two significant differences that were found might be a statistical fluke. We interpret these results as support for differentiating structure-mismatched items from regular MN-[ej] items because (1) the particular manner in which the significant differences have appeared is both consistent with our hypothesis and rather unlikely to have happened by chance, and (2) no significant difference was observed that runs counter to our hypothesis. Furthermore, finding statistically significant differences only in the reading task and not in the translation task suggests that the 'lagging' (structure-mismatched) items may be catching up to the 'leading' (regular MN-[ej]) items in showing the sound change, and that they are doing so first in the task associated with more crosslinguistic influence (translation). Such a pattern is compatible with our hypothesis that the merger reversal is attributable to crosslinguistic influence.

Consequently, while issuing the caveat that Study 2 was based on much fewer items than Study 1, we consider the pattern of results in this study to converge with the results of Study 1 in providing general support for our predictions.

4.3. STUDY 3: DISSIMILARITY IN ONSET. The goal of Study 3 was to examine how similarity in syllable onset conditions the sound change under investigation. Thus, Study 3 examined another subset of MN-[ej] items, which are maximally similar to their Mandarin counterpart with respect to syllable structure (i.e. CV in Shanghainese and CVX in Mandarin) but more dissimilar with respect to onset. The critical items under investigation were four MN-[ej] items that contain a voiced onset /b/,[13] but whose counterparts in Mandarin (which has no voiced plosives) contain a voiceless onset /p/ (n = 2) or /p^h/ (n = 2). We refer to this subset as onset-mismatched MN-[ej] items (or 'onset-mismatched items' for short).

These onset-mismatched items were compared to regular (i.e. onset-matched) MN-[ej] items (n = 4) and MN-[aj] items (n = 4) with voiceless onsets (/p p^h t t^h/). Since all onset-mismatched items had the voiced onset /b/, levels of Onset (PHTH, PTB) were not balanced across lexical sets. The scarcity of appropriate test items also prevented usage frequency from being balanced among the items. With three levels of Block (1, 2, 3), two levels of Age (Young, Old), and 12 speakers in each age group, the dataset for each experiment in Study 3 included about 864 tokens (12 items×3 blocks×2 ages×12 speakers), with approximately 3% missing due to unrecognized words or corrupted sound files.

Following the general modeling procedure used in the previous two studies, a set of 10 LM and GLM models was constructed to compare vowel formants and diphthongization rates in onset-mismatched items vs. regular MN-[ej] items and MN-[aj] items, with onset-mismatched items set as the reference level of LexSet for easier comparison.

READING VS. TRANSLATION. The final models showed some significant effects of LexSet near the onset of the vowel in both experiments. Interestingly, the general pattern of these effects suggested that the vowel of onset-mismatched items started from a more centralized position in the F1-F2 space than the vowels of the other lexical sets (see Figure 6). This was most evident when comparing onset-mismatched items with MN-[aj] items, as MN-[aj] items tended to have lower F1Start (Reading: $\beta_{\text{MN-[aj]}} = -21.17$, $t = -1.91$, n.s.; no LexSet×Age interaction. Translation: $\beta_{\text{MN-[aj]}} = -32.66$, $t = -5.03$, $p_{MCMC} = .002$; no LexSet×Age interaction) and higher F2Start (Reading: $\beta_{\text{MN-[aj]}} = 50.17$, $t = 2.53$, $p_{MCMC} = .021$; no LexSet×Age interaction. Translation: $\beta_{\text{MN-[aj]}} = 44.37$, $t = 2.88$, $p_{MCMC} = .004$; no LexSet×Age interaction) than onset-mismatched items. The comparison of onset-mismatched items with regular MN-[ej] items did not yield statistically significant results (all $|t|$'s < 2, all p_{MCMC}'s > .1), but similar trends can be observed when comparing the average starting positions of onset-mismatched items and regular MN-[ej] items (in mel, onset-mismatched MN-[ej]: $\text{F1Start}_{mean} = 598$, $\text{F2Start}_{mean} = 1527$; regular MN-[ej]: $\text{F1Start}_{mean} = 591$, $\text{F2Start}_{mean} = 1571$).

<INSERT FIGURE 6>

Since the first vowel preceding a critical vowel was usually lower and more back, the relatively centralized quality of the critical vowel in onset-mismatched items is most likely due to greater vowel-to-vowel coarticulation in these items, which would follow from the shorter closure duration of the intervening voiced stop in onset-mismatched items (Shen et al. 1987). An anonymous reviewer suggested that the breathy phonatory setting associated with the Shanghainese ‘voiced’ stops may also be playing a role; however, since breathy phonation is typically found to have a backing and/or RAISING effect on vowel quality (Gordon & Ladefoged 2001, Lotto et al. 1997, Thurgood 2000), this is unable to account for the lowering effect observed in the majority of cases.

While the start of the vowel in onset-mismatched items was characterized by relative centralization, a different pattern emerged near the offset of the vowel (Figure 6). In reading, there were no significant differences in F1End or F2End between onset-mismatched items and the other lexical sets,[14] but the average ending position of onset-mismatched items was between that of MN-[aj] items and that of regular MN-[ej] items (in mel, onset-mismatched MN-[ej]: $F1End_{mean} = 578$, $F2End_{mean} = 1594$; MN-[aj]: $F1End_{mean} = 615$, $F2End_{mean} = 1576$; regular MN-[ej]: $F1End_{mean} = 560$, $F2End_{mean} = 1616$). In translation, by contrast, the ending position of onset-mismatched items was significantly higher and more front than that of MN-[aj] items for both age groups (F1End: $\beta_{\text{MN-[aj]}} = 77.97$, $t = 6.61$, $p_{MCMC} < .001$; $\beta_{\text{MN-[aj]:Old}} = -44.27$, $t = -3.19$, $p_{MCMC} = .001$. F2End: $\beta_{\text{MN-[aj]}} = -69.39$, $t = -6.44$, $p_{MCMC} < .001$; $\beta_{\text{MN-[aj]:Old}} = 48.08$, $t = 4.84$, $p_{MCMC} < .001$), but hardly distinguishable from that of regular MN-[ej] items (in mel, onset-mismatched MN-[ej]: $F1End_{mean} = 568$, $F2End_{mean} = 1651$; regular MN-[ej]: $F1End_{mean} = 565$, $F2End_{mean} = 1652$; all $|t|$'s ≤ 1.65, $p_{MCMC} \geq .1$).

Given our findings from Study 1, these results suggest that the vowel in onset-mismatched items, like the vowel in regular MN-[ej] items, tended to move from [e] (or more likely [ɛ]) toward [i], and the tendency was stronger in translation. Results from the GLM models on Diphthong confirmed that in both experiments, onset-mismatched MN-[ej] items had a comparable rate of diphthongization toward [i] as regular MN-[ej] items ($|z| \leq 1.82$, $p(|z|) \geq .69$), which was significantly higher than that of MN-[aj] items ($|z| > 4$, $p(|z|) < .001$). The differences in rates of diphthongization are summarized in Figure 7. As shown in Figure 7, cross-lexical set differences in diphthongization rates tended to be more pronounced in reading than in translation.

<INSERT FIGURE 7>

CONTROL FACTORS. Study 3 also revealed some control effects that were consistent with the findings of Studies 1 and 2. Female speakers consistently produced higher-frequency formants than male speakers; older speakers overall produced more peripheral vowels (lower F1 and higher F2) than younger speakers; and items with aspirated plosive onsets (/p^h t^h/) had higher F1 than items with unaspirated plosive onsets (/p t/). Furthermore, there was a slight tendency for high-frequency items to show a greater difference between onset-mismatched and regular MN-[ej] items in F1Start, especially in reading (Reading: $\beta_{\text{MN-[ej]\_regular}} = -20.86$, $t = -1.88$, n.s.;

$\beta_{\text{MN-[ej]\_regular:HighFrq}}$ = -27.66, t = -2.79, p_{MCMC} = .028. Translation: $\beta_{\text{MN-[ej]\_regular}}$ = -8.00, t = -1.24, n.s.; $\beta_{\text{MN-[ej]\_regular:HighFrq}}$ = -11.98, t = -2.07, n.s.).

SUMMARY. To sum up, Study 3 yielded similar findings as Study 2. Onset-mismatched MN-[ej] items have undergone a similar sound change from [ɛ] to [ej] (or [ɛj]) as regular MN-[ej] items. On the whole, onset-mismatched items (much like structure-mismatched items in Study 2) patterned closely with regular MN-[ej] items, showing a statistically significant difference from regular MN-[ej] items only on the gradient measure of F1Start in the reading experiment; however, the comparison of average formant values and diphthongization rates across lexical sets consistently suggested a lesser degree of phonetic drift in onset-mismatched items than in regular MN-[ej] items. Thus, as in Study 2, differences between mismatched and regular MN-[ej] items were reduced in accordance with degree of crosslinguistic influence, surviving (barely) in reading and disappearing entirely in translation. These results are again consistent with our hypothesis.

5. DISCUSSION. The main goal of this research was to link group-level sound change with individual-level language processing mechanisms. To do so, we examined an ongoing sound change in contemporary Shanghainese, with a focus on its acoustic status and psycholinguistic basis. Our experimental results confirmed that the [ɛ] vowel in one Shanghainese lexical set (MN-[ej] items) is drifting toward [e] (or even [ej]), thereby distinguishing itself from the [ɛ] vowel in other lexical sets, which is not drifting in the same manner. The differences between the innovative and conservative vowel variants can exceed 100 mel in F1 and 60 mel in F2, and as discussed in §2, the rise of the innovative variant is effectively reversing a previous (near-)merger of /e/ and /ɛ/. In addition to examining the phonetic details of this sound change, our results provided evidence that the change is due to contact with Mandarin, as shown in marked diphthongization in the new Shanghainese vowel and greater distinctiveness of the new vowel compared to the old vowel among younger speakers (who are more bilingual in Mandarin), in bilingual mode (when Mandarin is more activated), and in word forms that are more similar phonologically to Mandarin counterparts. These facts cannot be explained by the alternative hypothesis attributing the reversal to remnants of [e] from an incomplete merger. The goals and results of Studies 1–3 are summarized in Table 5.

<INSERT TABLE 5>

Thus, in this research we have provided empirical evidence for the occurrence of a CONTACT-INDUCED MERGER REVERSAL, a phenomenon that, to our knowledge, has not been documented before and challenges the prevailing understanding of phonological mergers and merger reversals. As discussed in §2.1, previous cases of merger reversal have been attributed either to incompleteness of the relevant merger or to some other trace of the original contrast (e.g. phonological cues, orthographic cues, continuous existence in a different variety of the same language). In the case of the present Shanghainese merger reversal, too, the merger of /e/ and /ɛ/ in the late 1980s was probably incomplete; crucially, however, the hypothesized mechanism of merger reversal via crosslinguistic influence at the lexical-phonetic level does not rely on the incomplete status of the merger. Therefore, it is reasonable to suspect that the Shanghainese vowel merger would have undergone a similar reversal via the cross-language path even if the merger had been complete. If language contact can be understood to constitute a 'linguistic means' of precipitating a sound change, this suggests that complete mergers can indeed be reversed by linguistic means—contra Garde's principle—although this claim needs to be tested empirically with a merger that is truly complete.

The fact that language contact has heretofore not been documented as a possible means of merger reversal suggests that contact-induced merger reversal is relatively uncommon. In our view, this is at least partly attributable to its highly contingent environment. Figure 8 illustrates a series of events in time that characterize a contact-induced merger reversal. At Time 1, language L1 has two sets of words (Set A and Set B) with contrastive pronunciations (X and Y, respectively). At Time 2, Set B is merged into Set A by changing pronunciation from Y to X, and the distinction between the two sets is thus lost in L1. After the merger, speakers of L1 become influenced by another language, L2, which has a set of words that are cognate with Set B and pronounced as Y, and the phonological similarity of X and Y leads to cross-language perceptual linkage between Set B in L1 (pronounced as X now) and their cognates in L2 (pronounced as Y) in the minds of L1-L2 bilinguals. At Time 3, Set B in L1 readopts the Y pronunciation via influence from L2, which in effect reestablishes the contrast between Set A and Set B in L1.

<INSERT FIGURE 8>

As shown in Figure 8, several conditions need to be met for such a merger reversal to occur. First, it is necessary for L2 to share large sets of cognate words with L1. Since the cross-language perceptual linkage is established at the lexical-phonetic level between a pair of cognate words, it stands to reason that the more cognate words are shared between L1 and L2, the more individual instances of perceptual linkage there can be and, hence, the more powerful an effect these linkages can collectively exert on the L1 phonology. Second, Set B must have a pronunciation similar to Y in L2, making it possible for the Y pronunciation to be revitalized in L1. Third, the two pronunciations, X and Y, must be phonetically similar in order for cross-language perceptual linkage to be established after Time 2 (when Set B is pronounced with X in L1 and Y in L2). Finally, there also needs to be constant influence of L2 on L1 after (but not before) the merger of X and Y in L1. Given these conditions, a close relationship between L1 and L2 is most likely crucial, and the relative timing of the merger and crosslinguistic influence is critical, too. At this point, it is not clear whether it is also crucial that the contact language, L2, preserves the contrast between Set A (assuming that cognates of Set A also exist in L2) and Set B; it would be interesting to test for crosslinguistic influence in a situation where Set A and Set B have been merged to X in L1 and to Y in L2, if such a situation can be identified.

In contrast to previous studies (see §2.2), the older group in the current study did not show a clear merger across the three lexical sets as one might expect for this generation (speakers born in the 1940s and 1950s), but instead they differentiated the MN-[ej] lexical set from the other two, much like the younger generation. Assuming that Xu and Tang's (1988) findings faithfully reflect the production of the older generation when they were tested in the late 1970s and early 1980s (i.e. when they were in their 20s and 30s, which is around the time when the Chinese national language policy promoting Mandarin went into full effect), we believe that the Mandarin influence evident in this generation's production in 2011–2012 probably reflects their having been influenced over the intervening 30 years by the changing linguistic landscape in Shanghai brought about by the national language policy. Interestingly, if this is the case, that would make this an example of L1 phonetic change in the native language environment, as opposed to L1 phonetic change (attrition) in a foreign language environment (e.g. de Leeuw et al. 2010, Mayr et al. 2012). In other words, these individuals' L1 would have changed because of their environment changing around them, rather than them moving to a different environment.

In addition to the implications for the study of historical sound change, the current findings also have implications for theoretical models of bilingual speech. To begin with, our results show that for the Shanghainese merger reversal to have occurred via language contact, perceptual linkage between Shanghainese [ɛ] and Mandarin [ej] must have occurred at the lexical-phonetic level (see Figure 1), not simply at the general phonetic (namely, allophonic) level, as the [ɛ] vowel remains unchanged in some lexical items. This is consistent with our argument that contact-induced merger reversal is most likely to be found between languages that have a large number of cognates, which make it possible to establish cross-language linkage at the level of word forms (see Paterson 2011 for a different example of the effects of cognation on phonetic properties of speech production in bilinguals).

Our results suggested not only that crosslinguistic influence arose from perceptual linkages at the lexical-phonetic level, but also that the degree of crosslinguistic influence between similar segments in perceptually linked lexical items was modulated by the phonological similarity of the items in aspects other than the critical segment. Although the difference between Shanghainese [ɛ] and Mandarin [ej] was small enough to allow for perceptual linkage between Shanghainese and Mandarin cognate lexical items, the differences between Shanghainese [ɛ] and Mandarin [an] and [aj] were not; accordingly, only MN-[ej] items showed the sound change described in this paper. What is particularly noteworthy about our findings is they suggest that the similarity of other parts of the word form besides the segment undergoing change may influence the strength of the cross-language linkage that provides the basis for the change. As shown in our experimental results, if a Shanghainese lexical item and its Mandarin counterpart differed to a greater extent than the maximally parallel cognate pairs—with respect to either the presence of a medial approximant before the vowel or the voicing of the onset consonant—the amount of phonetic drift in the Shanghainese [ɛ] vowel tended to be smaller, ostensibly due to a relatively weaker cross-language linkage. This kind of variation among lexical sets that all contain the segment undergoing change suggests that a level of crosslinguistic analysis focusing on segment-to-segment relationships between an L1 and an L2 is inadequate for explaining the full range of data. Our findings argue instead for basing segmental predictions in bilingual speech research not only on segment-level correspondences, but also on comparisons at a higher level—namely, the word form—because information in the lexical context of a segment can affect how that segment may influence and be influenced by a similar segment in another language.

Finally, our results suggest that the bilingual language mode—that is, the simultaneous activation of the two languages in contact—plays an essential role in contact-induced sound change. We observed more crosslinguistic influence in the bilingual translation experiment than in the monolingual sentence-reading experiment, showing that language mode in the bilingual speaker can be manipulated in a laboratory environment (where there is no interlocutor in the experimental task and no change in previous language experience) and lead to differences in linguistic behavior (see also Antoniou et al. 2011, 2012). Brief exposure to auditory tokens from the L2 was enough to cause Shanghainese-Mandarin bilinguals' L1 production to drift toward the L2, indicating that cross-language priming is conducive to convergent phonetic drift. This leads us to speculate that common bilingual activities such as code-switching, during which both languages are activated, are important environments for initiating and accelerating crosslinguistic influence and contact-induced sound change. Previous literature has found evidence for crosslinguistic phonetic convergence during fluent code-switching (Bullock et al. 2006, Toribio et al. 2005) and other language switching tasks (Goldrick et al. 2014, Olson 2013), but the extent to which these kinds of short-term effects observed in the laboratory have long-term impact on a bilingual's phonological systems remains to be investigated.

Based on the discussion above, we can lay out a general picture of the process of contact-induced sound change as follows. First, cross-language linkage is established between similar sounds across languages through equivalence classification. As the SLM predicts, the linkage is bidirectional, and the strength of the linkage depends on the level of phonological similarity of the linked sounds, the level of phonological similarity of the lexical contexts in which they are embedded, and the speaker's experience in each language (e.g. in our study, the older speakers had much less experience with Mandarin than the younger speakers, and therefore they showed a reduced amount of Mandarin influence on their Shanghainese). After the linkage is established, its strength can further fluctuate with language mode. When both languages are highly activated, the cross-language channels are strengthened, causing more phonetic convergence between sounds that are perceptually linked across languages. Given enough time and large enough bilingual populations, such individual-level cross-language convergence may eventually develop into a group-level sound change.

6. CONCLUSION. If the locus of language contact is indeed the bilingual mind, one would expect bilingual language systems to be the birthplace for many contact-induced language changes. Consequently, the investigation of such systems should be an essential step in understanding contact-related linguistic phenomena. The contribution of this study is in showing how individual-centered inquiry can shed light on contact phenomena evident within a speech community. In particular, the dynamics of language interaction within the bilingual mind may help to explain constraints on the time course and generalizability of language changes stemming from a specific crosslinguistic similarity. Our findings suggest, for example, that the strength of L1-L2 perceptual linkages in the bilingual mental lexicon (based on overall, not strictly segmental, crosslinguistic phonological similarity) is positively correlated with the occurrence of contact-induced change.

Our findings also demonstrate the variability and flexibility of bilingual language systems, which point out several avenues for further research. One issue regards the extent to which bilingual activities may give rise to long-term adjustments in each language system. As shown in the current study, crosslinguistic influence was enhanced when both languages were activated, but how long such effects may last after an activity directly engaging both languages (such as code-switching) is still unclear. The answer to this question may lie in the comparison of diverse types of bilinguals with systematically different kinds of bilingual experience (e.g. late-onset L2 learners, heritage speakers, long-term residents in bilingual societies). Another issue regards the propagation of contact-induced change specific to one part of the lexicon to the lexicon/language at large. Will the sound change observed in the MN-[ej] lexical set spread to other Shanghainese word types (in particular, those without Mandarin cognates)? If so, how will it spread? A better understanding of these and related issues will contribute to a more comprehensive theory of contact-induced language change.

REFERENCES

ANTONIOU, MARK; CATHERINE T. BEST; MICHAEL D. TYLER; and CHRISTIAN KROOS. 2011. Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics* 39.558–70.

ANTONIOU, MARK; MICHAEL D. TYLER; and CATHERINE T BEST. 2012. Two ways to listen: Do bilinguals perceive stop-voicing differently according to language mode? *Journal of Phonetics* 40.582–94.

BAAYEN, R. HARALD; DOUGLAS J. DAVIDSON; and DOUGLAS M. BATES. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59.390–412.

BARANOWSKI, MACIEJ. 2007. *Phonological variation and change in the dialect of Charleston, South Carolina*. Durham: Duke University Press.

BATES, DOUGLAS; MARTIN MAECHLER; and BEN BOLKER. 2011. *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-39. Online: http://cran.r-project.org/package=lme4.

BEST, CATHERINE T., and MICHAEL D. TYLER. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. *Language experience in second language speech learning: In honor of James Emil Flege*, ed. by Ocke-Schwen Bohn and Murray J. Munro, 13–34. Amsterdam: John Benjamins.

BOERSMA, PAUL, and DAVID WEENINK. 2011. Praat: Doing phonetics by computer. Version 5.3. Online: http://www.praat.org.

BULLOCK, BARBARA E.; ALMEIDA JACQUELINE TORIBIO; VERÓNICA GONZÁLEZ; and AMANDA DALOLA. 2006. Language dominance and performance outcomes in bilingual pronunciation. *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference*, ed. by Mary Grantham O'Brien, Christine Shea, and John Archibald, 9–16. Somerville: Cascadilla Press.

CAO, JIANFEN, and IAN MADDIESON. 1992. An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics* 20.139–60.

CARAMAZZA, ALFONSO; ALESSANDRO LAUDANNA; and CRISTINA ROMANI. 1988. Lexical access and inflectional morphology. *Cognition* 28.297–332.

CHANG, CHARLES B. 2010. *First language phonetic drift during second language acquisition*. Berkeley: UC Berkeley dissertation.

CHANG, CHARLES B. 2012. Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40.249–68.

CHAO, YUEN REN. 1928. *Xiandai wuyu yanjiu* [Analysis of modern Wu dialect]. Beijing: Science Press.

CHEN, YIYA. 2008. The acoustic realization of vowels of Shanghai Chinese. *Journal of Phonetics* 36.629–48.

CHEN, ZHONGMIN. 1995. Shanghai shiqu hua yuyin yibai duo nian lai de yanbian [Phonetic variation of the Shanghai dialect in urban districts in the past 100 years]. *Wuyu he minyu de bijiao yanjiu* [Comparative studies among Wu and Min dialects], ed. by Tsu-Lin Mei, 18–31. Shanghai: Shanghai Education Press.

CLOPPER, CYNTHIA G. 2013. Modeling multi-level factors using linear mixed effects. *Proceedings of Meetings on Acoustics* 19.060028.

CLYNE, MICHAEL G. 1972. *Perspectives on language contact*. Melbourne: Hawthorne Press.

DE LEEUW, ESTHER; MONIKA S. SCHMID; and INEKE MENNEN. 2010. The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition* 13.33–40.

DUANMU, SAN. 1990. *A formal study of syllable, tone, stress and domain in Chinese languages*. Cambridge: MIT dissertation.

DUANMU, SAN. 2007. *The phonology of standard Chinese* (2nd edition). Oxford: Oxford University Press.

DUANMU, SAN. 2014. Syllable structure and stress. *The handbook of Chinese linguistics*, ed. by C.-T. James Huang, Yen-Hui Audrey Li, and Andrew Simpson, 422–42. Chichester: Wiley-Blackwell.

EDKINS, JOSEPH. 1868. *A grammar of colloquial Chinese, as exhibited in the Shanghai dialect* (2nd edition). Shanghai: Presbyterian Mission Press.

FERGUSON, CHARLES A., and CAROL B. FARWELL. 1975. Words and sounds in early language acquisition. *Language* 51.419–39.

FLEGE, JAMES EMIL. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15.47–65.

FLEGE, JAMES EMIL. 1995. Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, ed. by Winifred Strange, 233–72. Baltimore: York Press.

FLEGE, JAMES EMIL. 1996. English vowel productions by Dutch talkers: More evidence for the "similar" vs "new" distinction. *Second-language speech: Structure and process*, ed. by Allan R. James and Jonathan Leather, 11–52. Berlin: Mouton de Gruyter.

FORSTER, JONATHAN C. 2008. DMDX. Version 3.2.6.3. Online: http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm.

GARDE, PAUL. 1961. Réflexions sur les différences phonétiques entre les langues slaves. *Word* 17.34–62.

GOLDRICK, MATTHEW; ELIN RUNNQVIST; and ALBERT COSTA. 2014. Language switching makes pronunciation less nativelike. *Psychological Science* 25.1–6.

GORDON, MATTHEW, and PETER LADEFOGED. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29.383–406.

GROSJEAN, FRANÇOIS. 2001. The bilingual's language modes. *One mind, two languages: Bilingual language processing*, ed. by Janet L. Nicol, 1–22. Malden: Blackwell Publishers.

GU, QIN. 2007. *The influence of language contact on the phonetic evolution of the dialects in Shanghai urban districts*. Shanghai: Shanghai Normal University dissertation.

GUY, GREGORY. 1990. The sociolinguistic types of language change. *Diachronica* 7.47–67.

HASSELMO, NILS. 1970. Code-switching and modes of speaking. *Texas studies in bilingualism*, ed. by Glenn G. Gilbert, 179–210. Berlin: Walter de Gruyter.

HSIAO, YUCHAU E. 2002. Glide, syllable and tone: An introduction. *Chinese Phonology* 11.1–6.

IHALAINEN, OSSI. 1994. The dialects of England since 1776. *English in Britain and overseas: Origins and development*, ed. by Robert Burchfield, 197–274. Cambridge: Cambridge University Press.

KARLGREN, BERNHARD. 1926. *Études sur la phonologie de chinoise* (Archives d'Études Orientales, t. XV, 1–2). Stockholm: Norstedt.

KOCHETOV, ALEXEI. 2006. The role of social factors in the dynamics of sound change: A case study of a Russian dialect. *Language Variation and Change* 18.99–119.

LABOV, WILLIAM. 1975. On the use of the present to explain the past. *Proceedings of the 11th International Congress of Linguists* (vol. 2), ed. by Luigi Heilmann, 825–51. Bologna: Mulino.

LABOV, WILLIAM. 1994. *Principles of linguistic change, Volume 1: Internal factors*. Malden: Blackwell Publishers.

LABOV, WILLIAM; MARK KAREN; and COREY MILLER. 1991. Near-mergers and the suspension of phonemic contrast. *Language Variation and Change* 3.33–74.

LAI, LI-FANG, and HUIJU HSU. 2013. Contact-induced sound change: Analysis of the alveolar lateral fricative in Yami. *Journal of Estonian and Finno-Ugric Linguistics* 4.31–50.

LINDBLOM, BJÖRN. 1990. Explaining phonetic variation: A sketch of the H and H theory. *Speech production and speech modeling*, ed. by William J. Hardcastle and Alain Marchal, 403–39. Dordrecht: Kluwer.

LOTTO, ANDREW J.; LORI L. HOLT; and KEITH R. KLUENDER. 1997. Effect of voice quality on perceived height of English vowels. *Phonetica* 54.76–93.

MAYR, ROBERT; SACHA PRICE; and INEKE MENNEN. 2012. First language attrition in the speech of Dutch–English bilinguals: The case of monozygotic twin sisters. *Bilingualism: Language and Cognition* 15.687–700.

METSALA, JAMIE L., and AMANDA C. WALLEY. 1998. Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. *Word recognition in beginning literacy*, ed. by Jamie L. Metsala and Linnea C. Ehri, 89–120. Mahwah: Lawrence Erlbaum Associates.

MICHELENA, LUIS. 1957. Las antiguas consonantes vascas. *Miscelánea homenaje a A. Martinet* (vol. 1), ed. by Diego Catalán, 113–57. La Laguna: Universidad de La Laguna.

MOON, SEUNG-JAE, and BJÖRN LINDBLOM. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America* 96.40–55.

NAGY, NAOMI. 1997. Modeling contact-induced language change. *University of Pennsylvania Working Papers in Linguistics* 4.399–418.

OLSON, DANIEL J. 2013. Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics* 41.407–20.

PATERSON, NATTALIA. 2011. *Interactions in bilingual speech processing*. Evanston: Northwestern University dissertation.

QIAN, NAIRONG. 2003. *Shanghai yuyan fazhan shi* [History of the Shanghainese language]. Shanghai: Shanghai People's Publishing House.

QIAN, NAIRONG. 2007. *Shanghai fangyan* [Shanghai dialect]. Shanghai: Wenhui Press.

R CORE TEAM. 2012. *R: A language and environment for statistical computing*. Vienna, Austria. ISBN: 3-900051-07-0. Online: http://www.R-project.org.

RATTE, ALEXANDER TAKENOBU. 2011. *Contact-induced phonological change in Taiwanese*. Columbus: Ohio State University MA thesis.

REN, NIANQI. 1987. *An acoustic study of Shanghai stops*. Storrs: University of Connecticut dissertation.

SANCIER, MICHELE L., and CAROL A. FOWLER. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 27.421–36.

SANKOFF, GILLIAN. 2002. Linguistic outcomes of language contact. *The handbook of language variation and change* (vol. 1), ed. by Jack K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 638–68. Malden: Blackwell Publishing.

SAPIR, EDWARD. 1921. *Language: An introduction to the study of speech.* New York: Harcourt Brace.

SHEN, TONG. 1981. Shanghaihua laopai xinpai de chabie [Differences between old and new varieties of Shanghainese]. *Fangyan* [Dialect] 1981-4.275–83.

SHEN, ZHONGWEI; CHARLES WOOTERS; and WILLIAM S.-Y. WANG. 1987. Closure duration in the classification of stops: A statistical analysis. *Ohio State University Working Papers in Linguistics* 35.197–209.

SHI, RUJIE, and JIANPING JIANG. 1987. Shanghai shiqu zhongnianren yuyin gongshi chayi de wubai ren diaocha [A survey of phonetic variation among 500 middle-aged speakers in urban districts of Shanghai]. *Yuyan Yanjiu Jikan* [Collected Papers on Language Studies] 1.271–96. Shanghai: Fudan University Press.

STEVENS, STANLEY SMITH; JOHN VOLKMANN; and EDWIN B. NEWMAN. 1937. A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America* 8.185–90.

SVANTESSON, JAN-OLOF. 1989. Shanghai vowels. *Lund Working Papers in Linguistics* 35.191–202.

THURGOOD, GRAHAM. 2000. Voice quality differences and the origin of diphthongs. *Berkeley Linguistics Society* 26.295–303.

TORIBIO, ALMEIDA JACQUELINE; BARBARA E. BULLOCK; CHRISTOPHER G. BOTERO; and KRISTOPHER ALLEN DAVIS. 2005. Perseverative phonetic effects in bilingual code-switching. *Theoretical and experimental approaches to Romance linguistics: Selected papers from the 34th Linguistic Symposium on Romance Languages*, ed. by Randall Gess and Edward J. Rubin, 291–306. Amsterdam: John Benjamins.

TRASK, R. L. 2000. *The dictionary of historical and comparative linguistics*. Edinburgh: Edinburgh University Press.

VAN COETSEM, FRANS. 1988. *Loan phonology and the two transfer types in language contact*. Providence: Foris Publications.

WEINREICH, URIEL. 1968. *Languages in contact: Findings and problems*. The Hague: Mouton de Gruyter.

WEINREICH, URIEL; WILLIAM LABOV; and MARVIN I. HERZOG. 1968. *Empirical foundations for a theory of language change*. Austin: University of Texas Press.

WINFORD, DONALD. 2005. Contact-induced changes: Classification and processes. *Diachronica* 22.373–427.

WU, ZONGJI (ed.) 1986. *Hanyu putonghua danyinjie yutu ce* [The spectrographic album of monosyllables of standard Chinese]. Beijing: China Social Sciences Press.

XU, BAOHUA, and ZHENZHU TANG (eds.) 1988. *Shanghai shiqu fangyan zhi* [Documentation of the Shanghai dialect in urban districts]. Shanghai: Shanghai Education Press.

XU, BAOHUA; ZHENZHU TANG; and ZHIXIANG TANG. 1986. Shanghai shiqu shaonian suo shiyong de shanghaihua yuyin [Phonetics of the Shanghainese spoken by teenagers]. *Yuwen Luncong* [Collections of Chinese Studies] 3.112–20. Shanghai: Shanghai Education Press.

[1] Any scholar familiar with varieties of Chinese would agree that the line between 'dialect' and 'language' is not often clear. A Chinese 'dialect' may, for example, differ to such a degree from the standard language that the two are not mutually intelligible. In fact, use of the term 'dialect' or 'language' is often based on considerations external to linguistics. As such, we use the two terms interchangeably when referring to varieties of Chinese in this paper.

[2] Given the contrast between two categories in the mid back region of the vowel space, it is conceivable that a pressure towards systemic symmetry, as well as the occurrence of a rounded category in the mid front region, could enhance the likelihood of a merged /ɛ/ 'splitting' into two mid front unrounded vowels. However, this is not likely to be the cause of the observed change, because if it were we would expect the lexical sets examined to pattern equivalently (since they all contain the same vowel eligible for splitting), yet they do not (see §4).

[3] Classification into lexical sets was done only to distinguish items according to their behavior in the series of changes involving [e] and [ɛ] in Shanghainese. Consequently, the notion of 'lexical set' should not be confused with the notion of 'rhyme group' used in traditional Chinese philology.

[4] Starting from Stage II, the vowel in MN-[aj] items was documented in some studies as halfway between [e] and [ɛ], i.e. [ɛ̝] (e.g. Shi & Jiang 1987); however, there is no evidence that this subtle difference from canonical [e] and [ɛ] was reliably perceptible or phonemic. Therefore, we follow the general convention and use the symbol [ɛ] for these cases.

[5] Given the published acoustic norms for Shanghainese and Mandarin vowels, it may seem that Shanghainese [ɛ] had even lower F1 than Mandarin [ej]. However, such a comparison should be taken with a grain of salt because the Mandarin norms in Wu 1986 were based on only two speakers (one for each sex) and there was considerable variability in syllable onset across these studies.

[6] Model trends reported in this paper have all been verified with alternative model analyses in which LexSet and Age were set to different reference levels.

[7] Like factors with treatment contrast coding, a factor with sum contrast coding is fitted with N-1 coefficients in a regression analysis, where N is the number of distinct levels of the factor. The i^{th} coefficient β_i represents the predicted difference between the i^{th} level of the factor (the

alphabetically i^th level by default) and the grand mean when other predictors are controlled. The coefficient associated with the N^th level of the factor can be computed by hand as the negative sum of β_1, β_2 … β_{N-1}, but its significance is not directly examined (Clopper 2013).

[8] After removing obviously nonsignificant fixed-effect predictors, we further explored the modeling of random effects by constructing alternative models with more complex random-effect structures (adding by-item and by-family random slopes, as well as by-speaker random intercepts and slopes). The more complex models generated similar results for the fixed effects as the models with simple random-effect structures (i.e. by-item and by-family intercepts only), but often showed signs of overparameterization (e.g. perfect correlation between two random effects) and less stable fit (e.g. singular convergence). Thus, for the sake of simplicity and reliability we report results only from the final models with simple random-effect structures.

[9] It is generally agreed in the analysis of Chinese syllable structure that the syllable-final glide /j/ and nasal /n/ are part of the rhyme but outside of the nucleus (see Duanmu 2014 for a review). Thus, the Shanghainese test items in Study 1 have a CV syllable structure while their Mandarin cognates have a CVX (where X = final, either glide or nasal) syllable structure. This is the highest level of crosslinguistic structural similarity that can be achieved for lexical items involved in the sound change at issue.

[10] The model of F2End showed a positive, but nonsignificant, main effect of LexSet = MN-[an] and a negative, significant interaction when LexSet = MN-[an] and Age = Old. However, an alternative analysis with Old as the reference level of Age indicated a null effect of LexSet = MN-[an] on F2End among older speakers ($\beta_{\text{MN-[an]}} = 3.67$, $t = 0.23$, n.s.), suggesting that the MN-[aj] vs. MN-[an] difference was not significant among older speakers, either.

[11] Whether the medial approximant in a Mandarin syllable (i.e. G in CGV and CGVX syllables) belongs in the onset, the rhyme, or a slot of its own is still under debate (see Hsiao 2002 and references therein). In this research, we remain agnostic as to the specific location of the medial /w/ and just consider structure-mismatched MN-[ej] items to differ in syllable structure from their Mandarin /Cwej/ counterparts to a greater degree than regular MN-[ej] items vis-a-vis their Mandarin /Cej/ counterparts. It should be noted, however, that if one were to follow the analysis of Duanmu (1990, 2007, 2014), wherein a medial G is analyzed as a secondary articulation on the initial C, our structure-mismatched MN-[ej] items would be ‘onset-mismatched’ instead,

---

such that the critical items in Study 2 and Study 3 would exemplify two different types of onset mismatch between Shanghainese and Mandarin cognates (i.e. mismatch in presence/absence of a secondary articulation in Study 2 and mismatch in onset voicing in Study 3). As pointed out by an anonymous reviewer, it is also possible to analyze the entire /wej/ sequence in structure-mismatched MN-[ej] items as a triphthongal V, which would render these syllables structurally the same as CV syllables. However, this proposal conflicts with current analyses of Mandarin syllable structure, which understand GVX as a hierarchical (i.e. not flat) structure with G further away than X from the vowel nucleus.

[12] An alternative model of F1End with Old as the reference level of Age showed that the effect of MN-[ej] was not significant among older speakers ($\beta_{\text{MN-[ej]}} = 5.83$, $t = 0.49$, n.s.).

[13] Despite the controversy regarding the voiced and breathy qualities of 'voiced' stop onsets in Shanghainese, most studies agree that this series of stops has short VOT similar to that of voiceless unaspirated stops (Cao & Maddieson 1992, Ren 1987).

[14] The model of F2End showed a negative, but nonsignificant, main effect of LexSet = MN-[aj] ($\beta_{\text{MN-[aj]}} = -33.39$, $t = -1.12$, n.s.) and a positive, significant interaction when LexSet = MN-[aj] and Age = Old ($\beta_{\text{MN-[aj]}} = 31.57$, $t = 3.13$, $p_{MCMC} = .002$). However, an alternative analysis with Old as the reference level of Age indicated that the effect of LexSet = MN-[aj] on F2End among older speakers was nonsignificant ($\beta_{\text{MN-[an]}} = -1.82$, $t = -0.06$, n.s.).

Lexical set	**Example**	**Mandarin rhyme**	**Shanghainese rhyme**			
			Stage I (1850–1920s)	Stage II (1920s–1960s)	Stage III (1970s–1980s)	Stage IV (after 2000)
MN-[aj]	来 ‘to come’ pinyin: *lái*	[aj]	[e]	[ɛ]	[ɛ]	[ɛ]
MN-[an]	兰 ‘orchid’ pinyin: *lán*	[an]	[ɛ]	[ɛ]	[ɛ]	[ɛ]
MN-[ej]	雷 ‘thunder’ pinyin: *léi*	[ej]	[e]	[e]	[ɛ]	[e]

TABLE 1. Historical development of the rhymes in the MN-[aj], MN-[an], and MN-[ej] lexical sets of Shanghainese.

Outcome variable	F1Start / F2Start / F1End / F2End / Diphthong
Random effects	by-item intercept by-family intercept
Fixed effects	LexSet×Age LexSet×Sex (LexSet×Frq)[a] Block Onset

TABLE 2. Initial model structure in Studies 1–3.

[a] LexSet×Frq was included in the initial models for Studies 1 and 3, but not in the models for Study 2, because Frq did not vary among the items in Study 2 (see §4.2 for more details).

Predictor	**Coding scheme**	**Levels in Study 1**	**Levels in Study 2**	**Levels in Study 3**
LexSet	Treatment	MN-[aj], MN-[an], MN-[ej]	structure-mismatched MN-[ej], MN-[aj], regular MN-[ej]	onset-mismatched MN-[ej], MN-[aj], regular MN-[ej]
Age	Treatment	Old, Young	Old, Young	Old, Young
Block	Orthogonal polynomial	1, 2, 3	1, 2, 3	1, 2, 3
Frq	Sum contrast	High, Low	--[a]	High, Low
Onset	Sum contrast	L, PHTH, PTB (with /p t/ only)	PHTH, PTB (with /p t/ only), S	PHTH, PTB
Sex	Sum contrast	F, M	F, M	F, M

TABLE 3. Properties of fixed-effect predictors in regression analyses in Studies 1–3. Reference levels are underlined. L = /l/ onset; PHTH = /p^h/ or /t^h/ onset; PT = /p/ or /t/ onset; S = /s/ onset.

[a] Frq was not a predictor in models for Study 2.

	Reading			Translation		
	MN-[aj]	MN-[an]	MN-[ej]	MN-[aj]	MN-[an]	MN-[ej]
Younger talkers	131	120	125	143	148	164
Older talkers	133	125	127	169	163	174

TABLE 4. Mean vowel duration (in ms) in each lexical set, by talker age and experimental task (Study 1).

	Study 1	Study 2	Study 3
Goal(s)	• confirm difference between MN-[ej] and other lexical sets • test for effects of age (degree of bilingualism) and task (language mode)	• examine influence of phonological (dis)similarity in syllable structure	• examine influence of phonological (dis)similarity in syllable onset
Results	• reading: MN-[ej] items drifting toward [ej] (effect of LexSet on F1End, F2End, Diphthong), with a greater degree of drift in younger talkers (LexSet x Age interaction) • translation: same effects, but more pronounced	• reading: drift toward [ej] also present in structure-mismatched items (effect of LexSet = MN-[aj] on F1End, F2End, Diphthong) but to a lesser degree than in regular MN-[ej] items (effect of LexSet = MN-[ej]_regular on F1End and Diphthong) • translation: similar effects, but structure-mismatched and regular MN-[ej] items pattern more closely	• reading: drift toward [ej] also present in onset-mismatched items (effect of LexSet = MN-[aj] on Diphthong), with traces of an intermediate degree of drift between MN-[aj] and regular MN-[ej] items • translation: similar effects, but onset-mismatched and regular MN-[ej] items pattern more closely

TABLE 5. Goals and results of Study 1, Study 2, and Study 3.

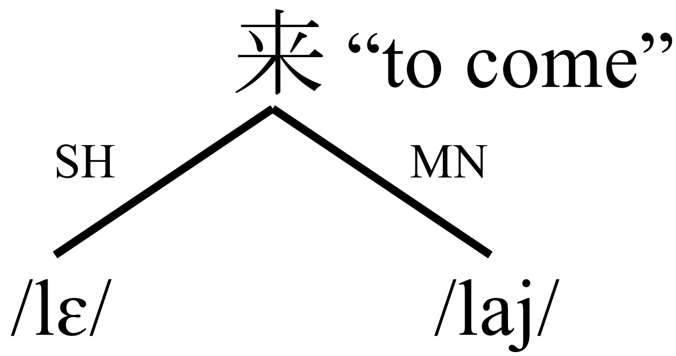

FIGURE 1a. No cross-language linkage in MN-[aj] items (e.g. 来 'to come').

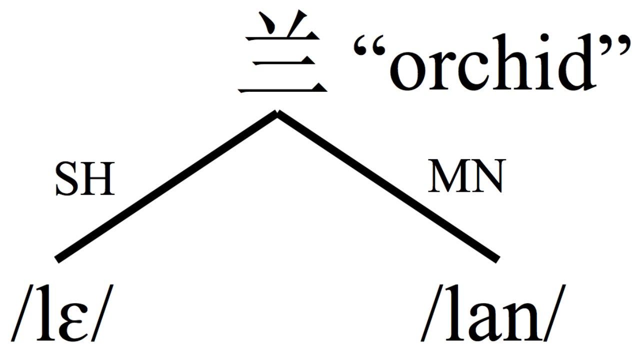

FIGURE 1b. No cross-language linkage in MN-[an] items (e.g. 兰 'orchid').

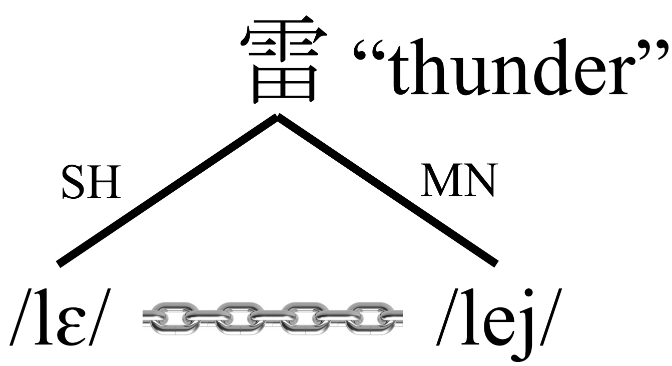

FIGURE 1c. Cross-language linkage in MN-[ej] items (e.g. 雷 'thunder').

FIGURE 1. Bilingual lexical representations in the three lexical sets. SH = Shanghainese, MN = Mandarin.

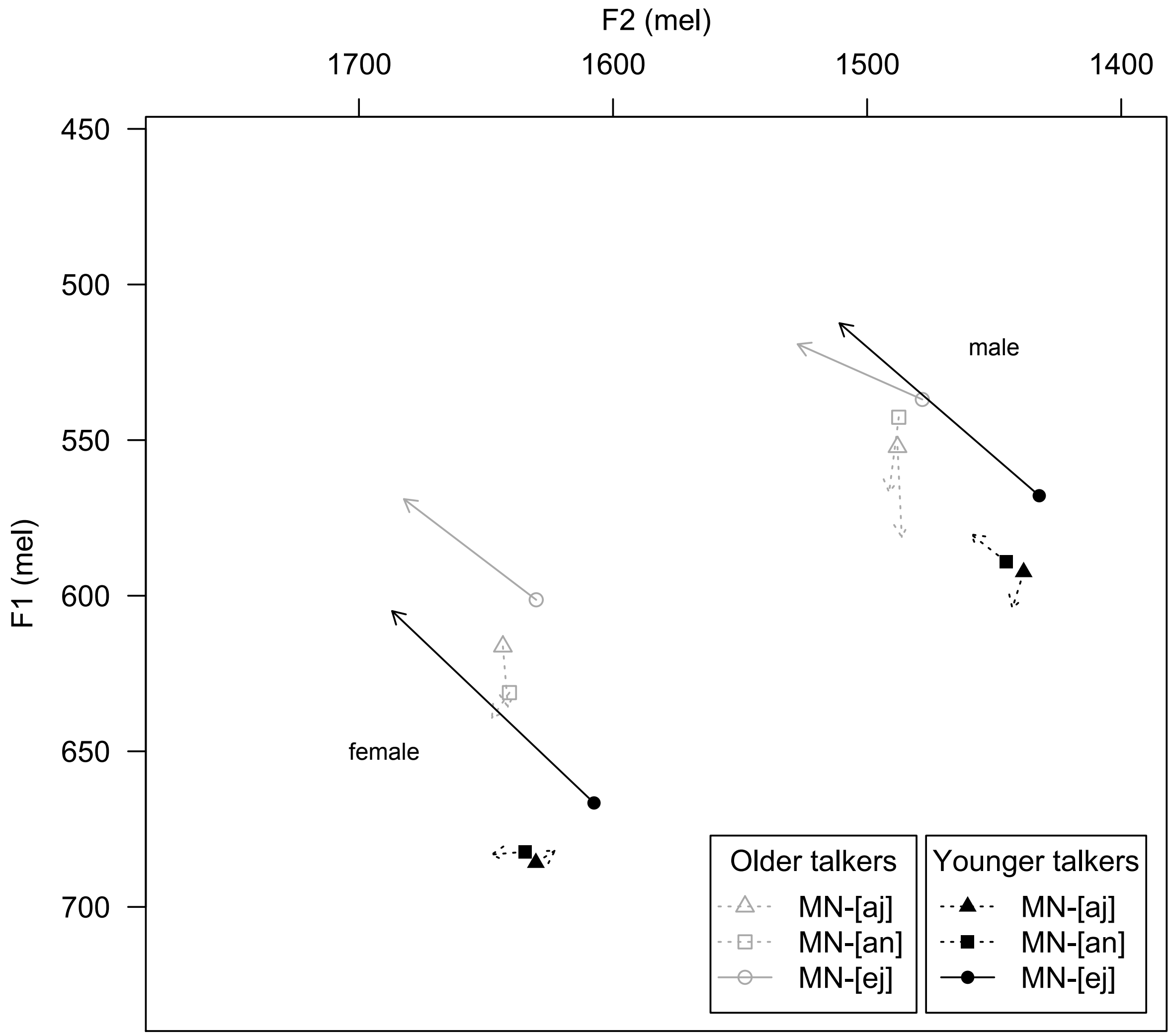

FIGURE 2a. Reading.

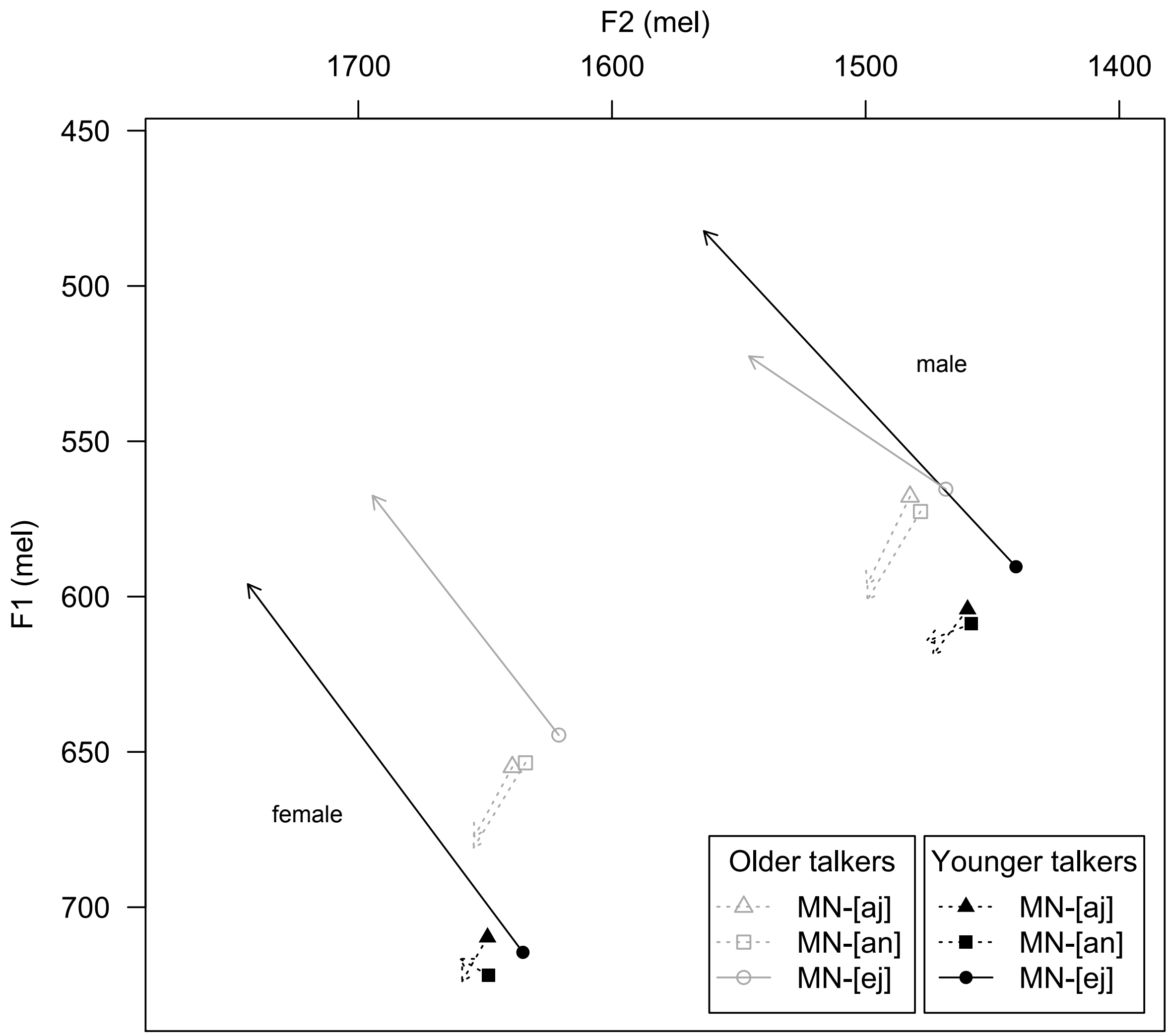

FIGURE 2b. Translation.

FIGURE 2. Mean formant trajectories in Study 1 in (a) reading and (b) translation, by LexSet, Age, and Sex.

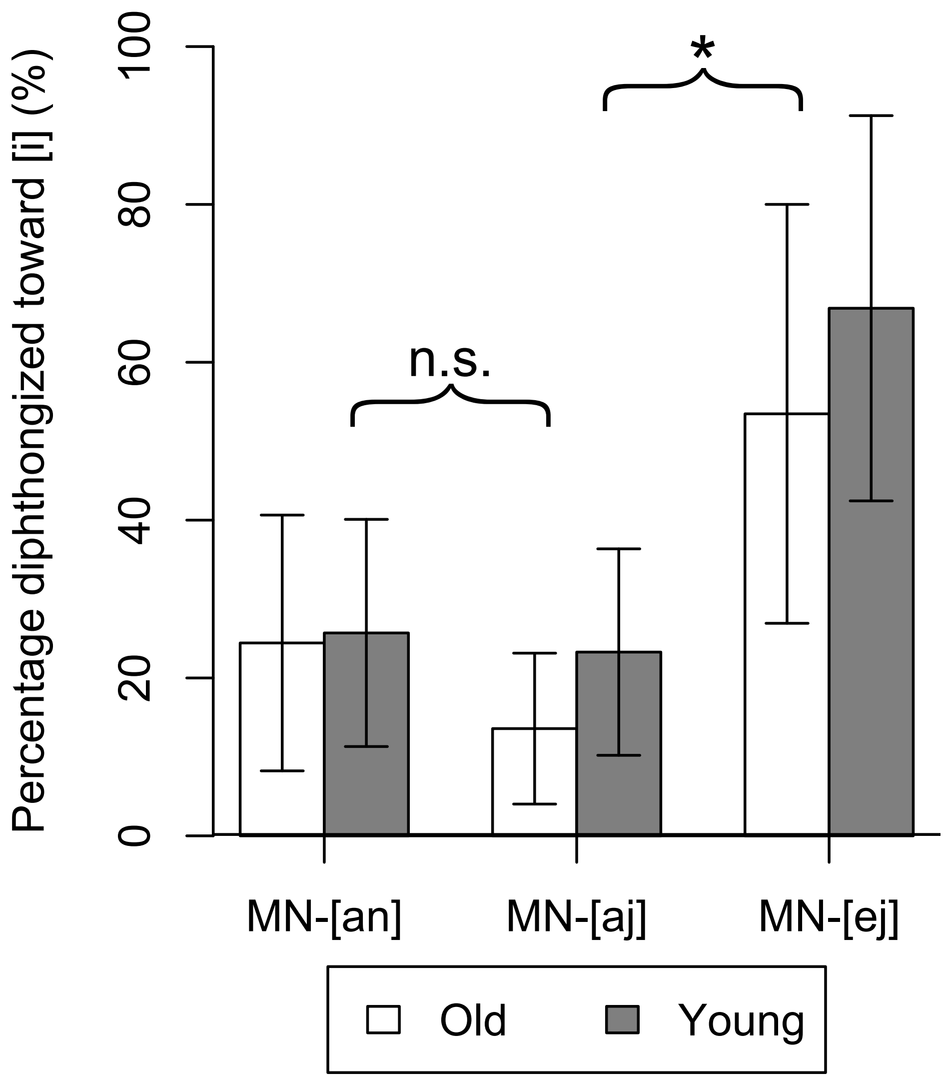

FIGURE 3a. Reading.

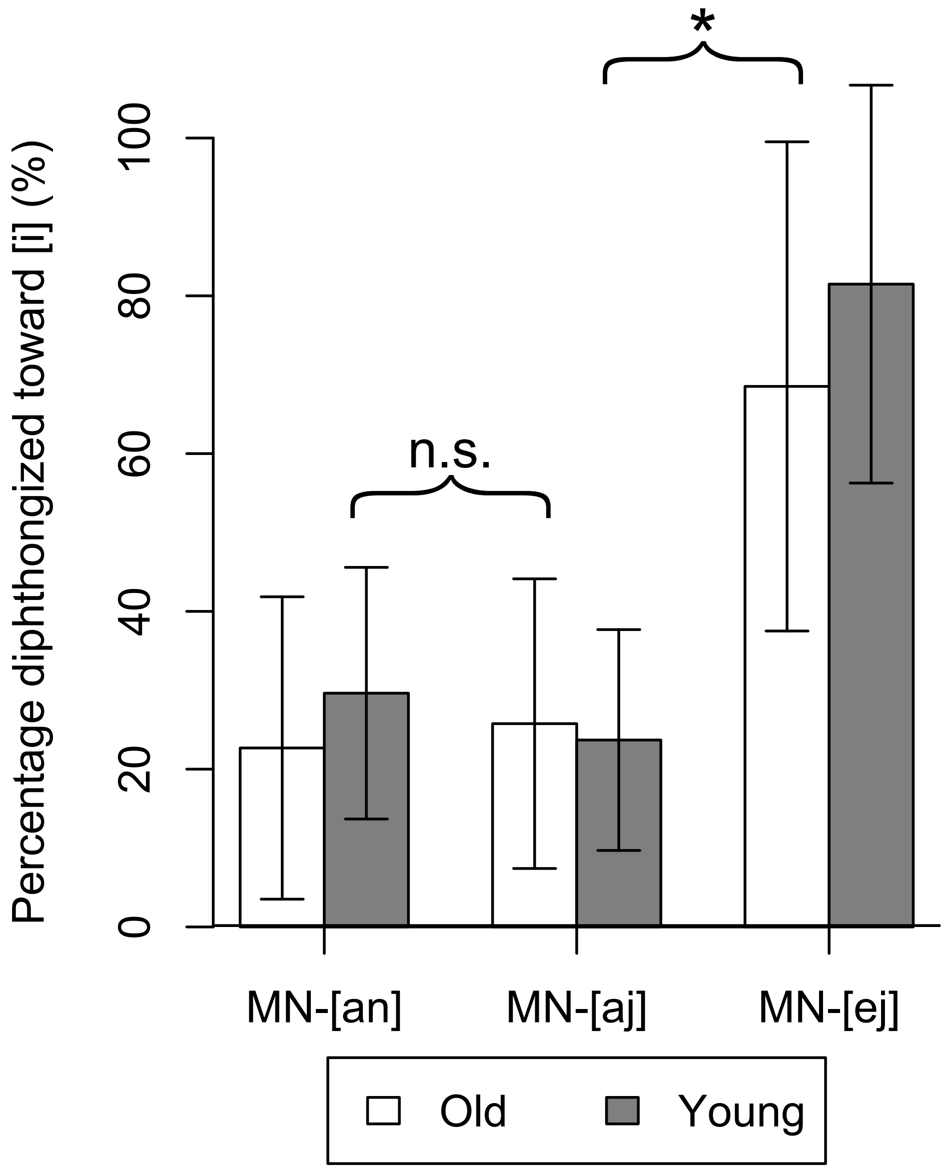

FIGURE 3b. Translation.

FIGURE 3. Mean percentage of vowels diphthongized toward [i] in Study 1 in (a) reading and (b) translation, by LexSet and Age. Error bars represent the range within one standard error of the mean over participants.

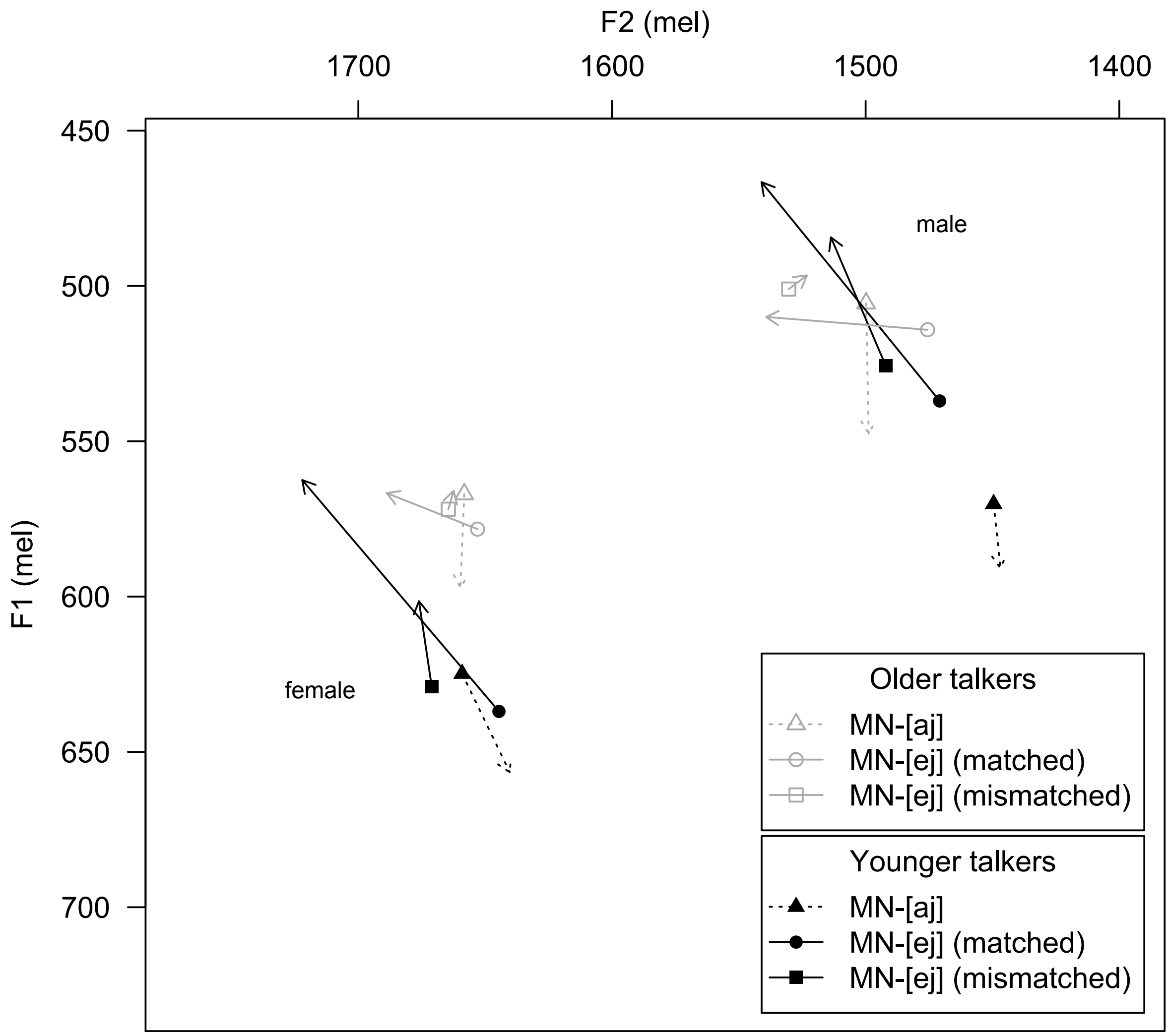

FIGURE 4a. Reading.

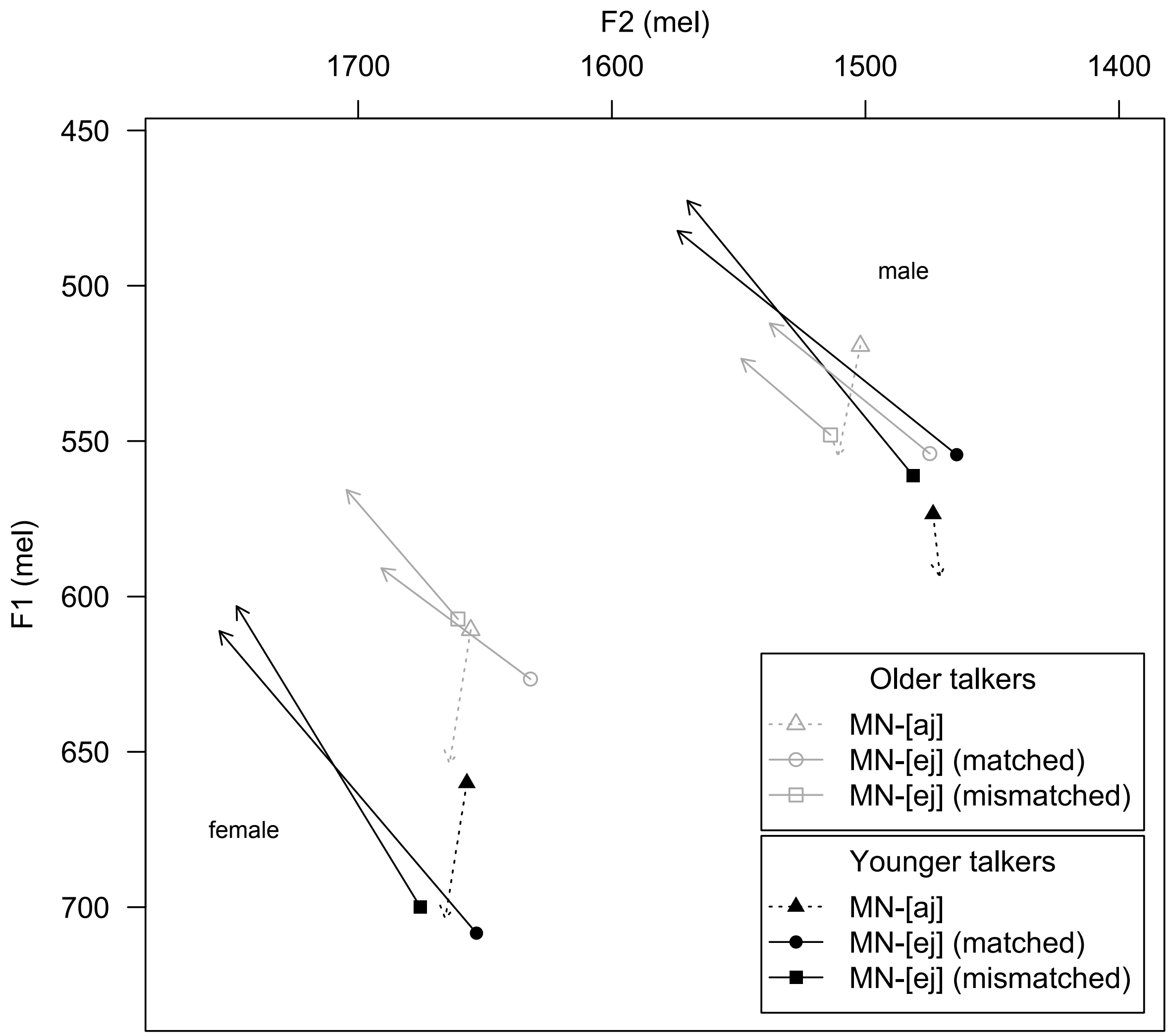

FIGURE 4b. Translation.

FIGURE 4. Mean formant trajectories in Study 2 in (a) reading and (b) translation, by LexSet, Age, and Sex.

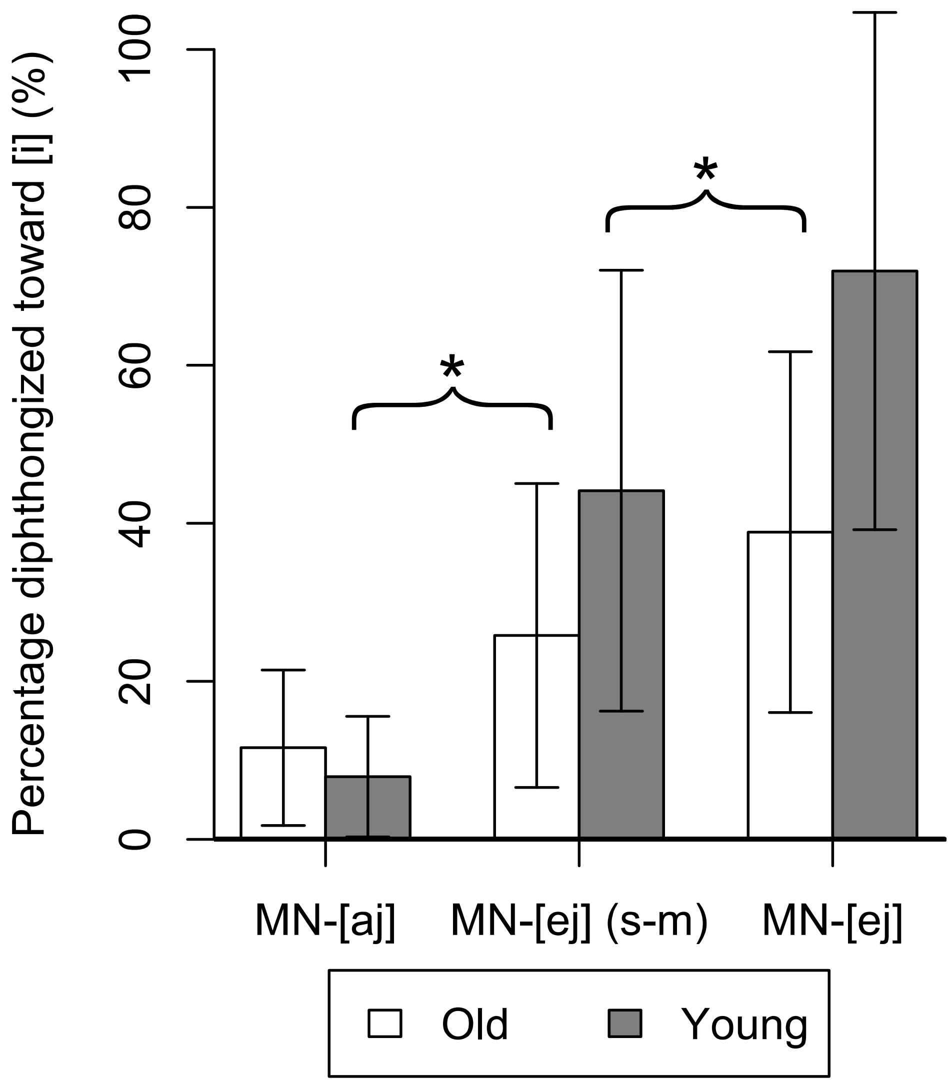

FIGURE 5a. Reading.

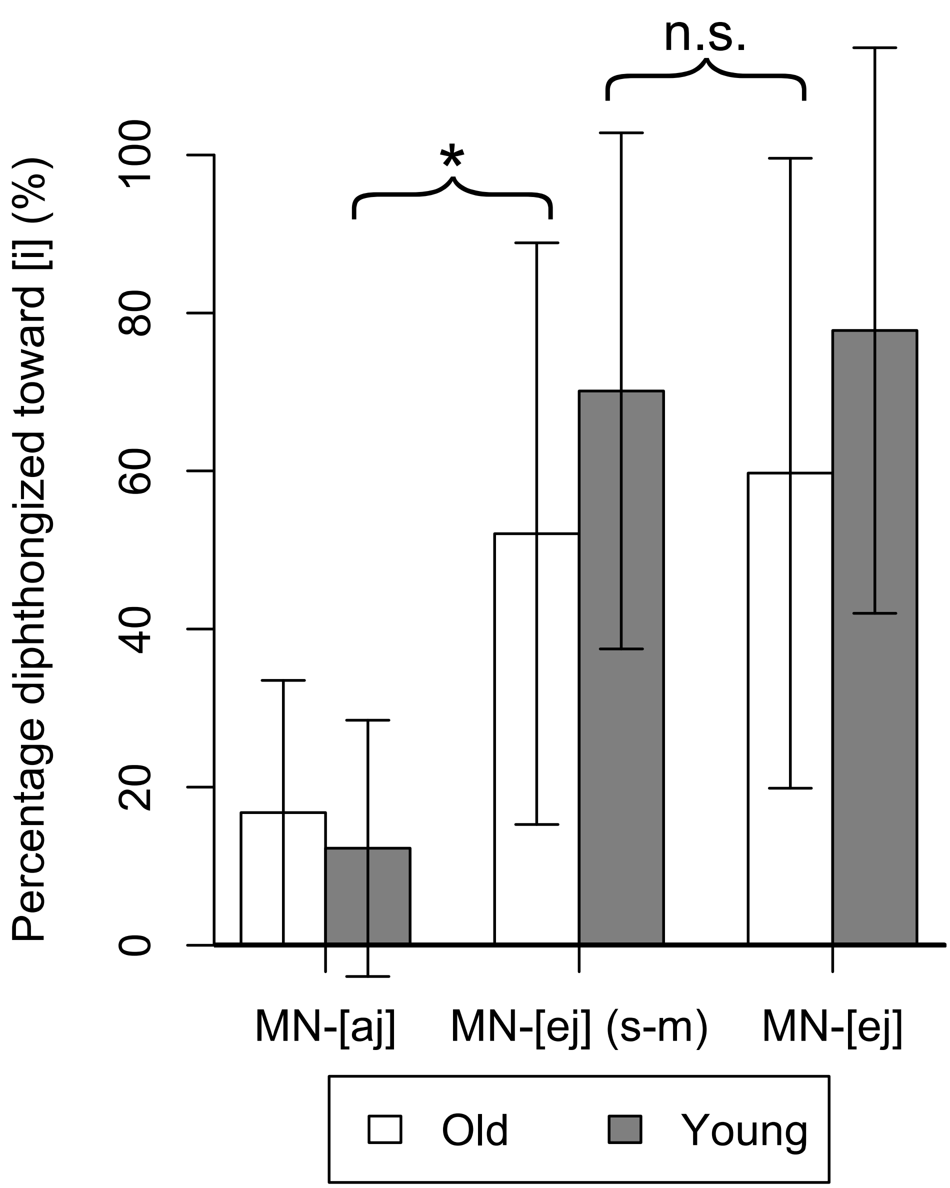

FIGURE 5b. Translation.

FIGURE 5. Mean percentage of vowels diphthongized toward [i] in Study 2 in (a) reading and (b) translation, by LexSet ("s-m" = structure-mismatched) and Age. Error bars represent the range within one standard error of the mean over participants.

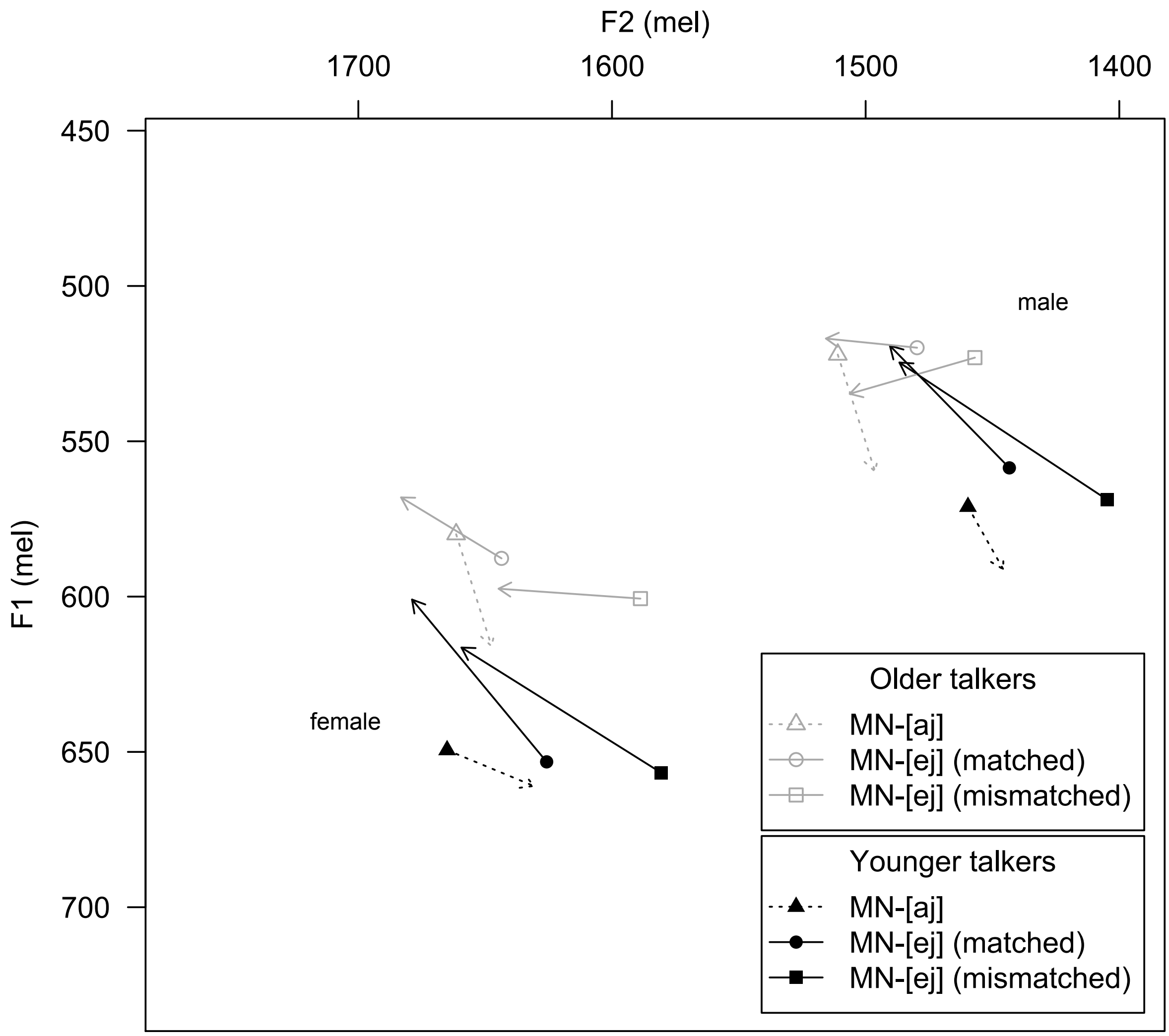

FIGURE 6a. Reading.

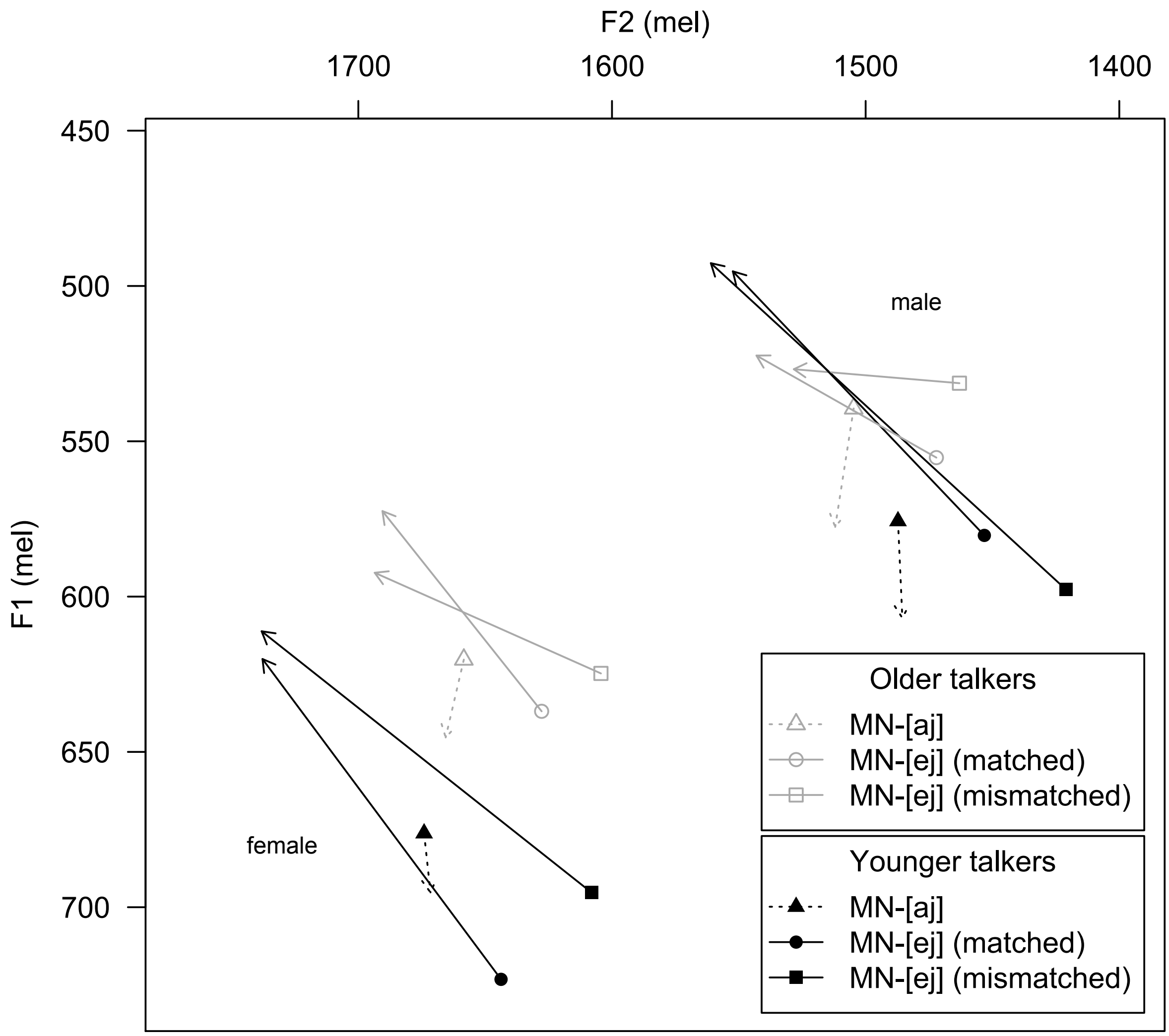

FIGURE 6b. Translation.

FIGURE 6. Mean formant trajectories in Study 3 in (a) reading and (b) translation, by LexSet, Age, and Sex.

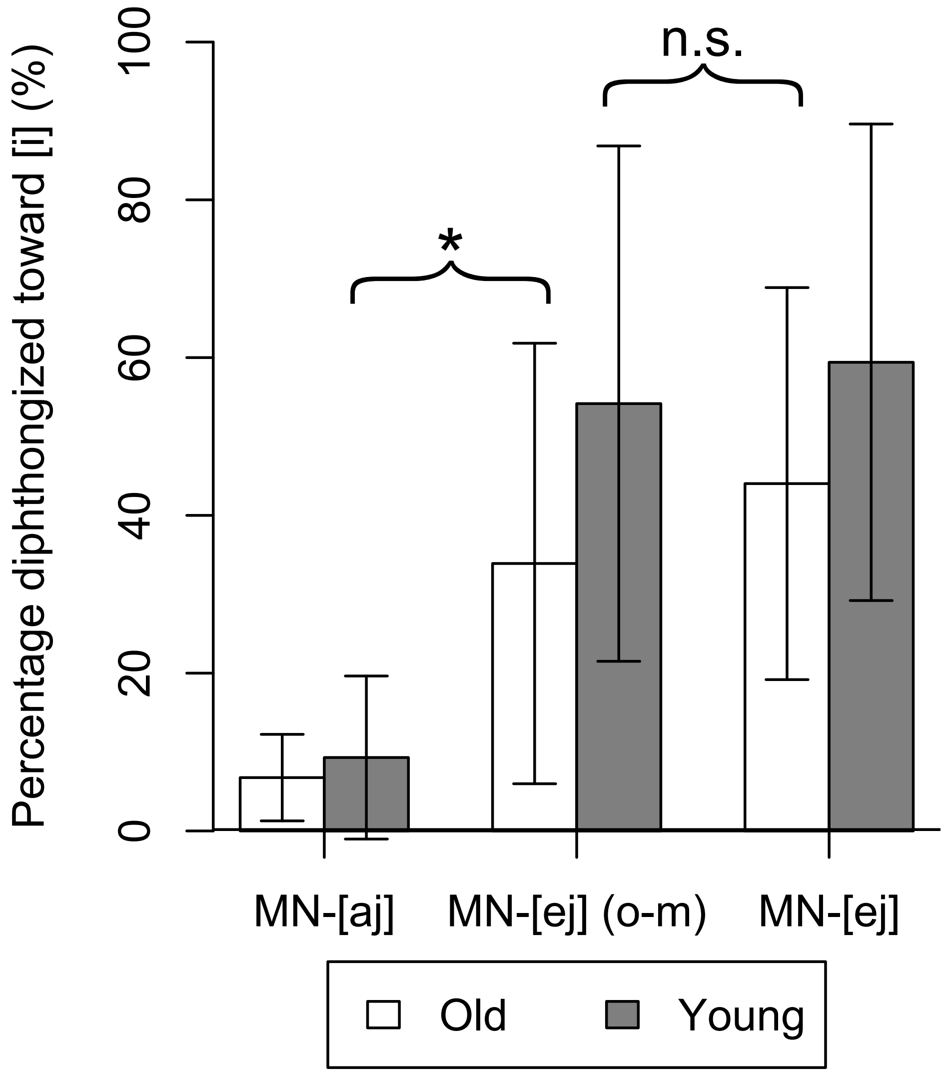

FIGURE 7a. Reading.

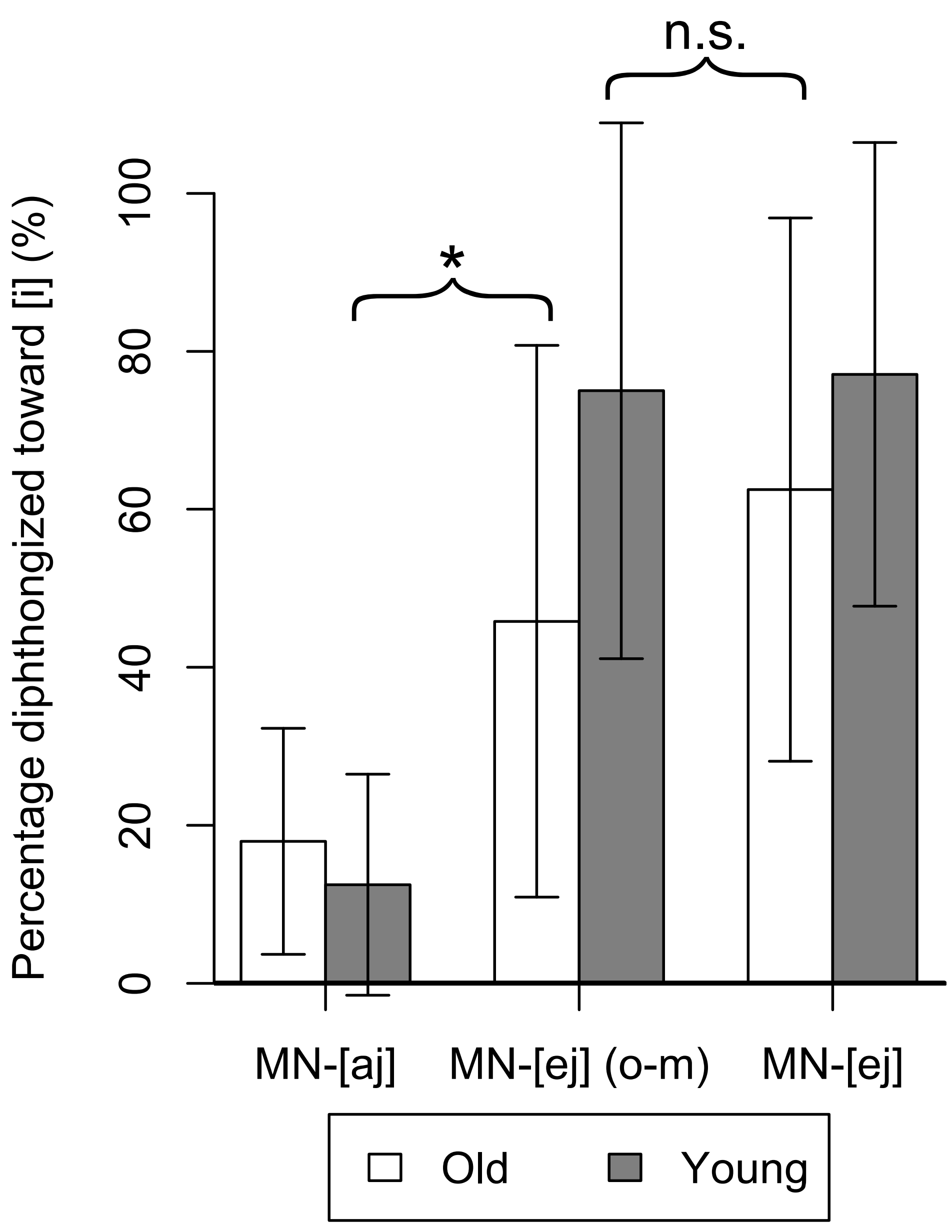

FIGURE 7b. Translation.

FIGURE 7. Mean percentage of vowels diphthongized toward [i] in Study 3 in (a) reading and (b) translation, by LexSet ("o-m" = onset-mismatched) and Age. Error bars represent the range within one standard error of the mean over participants.

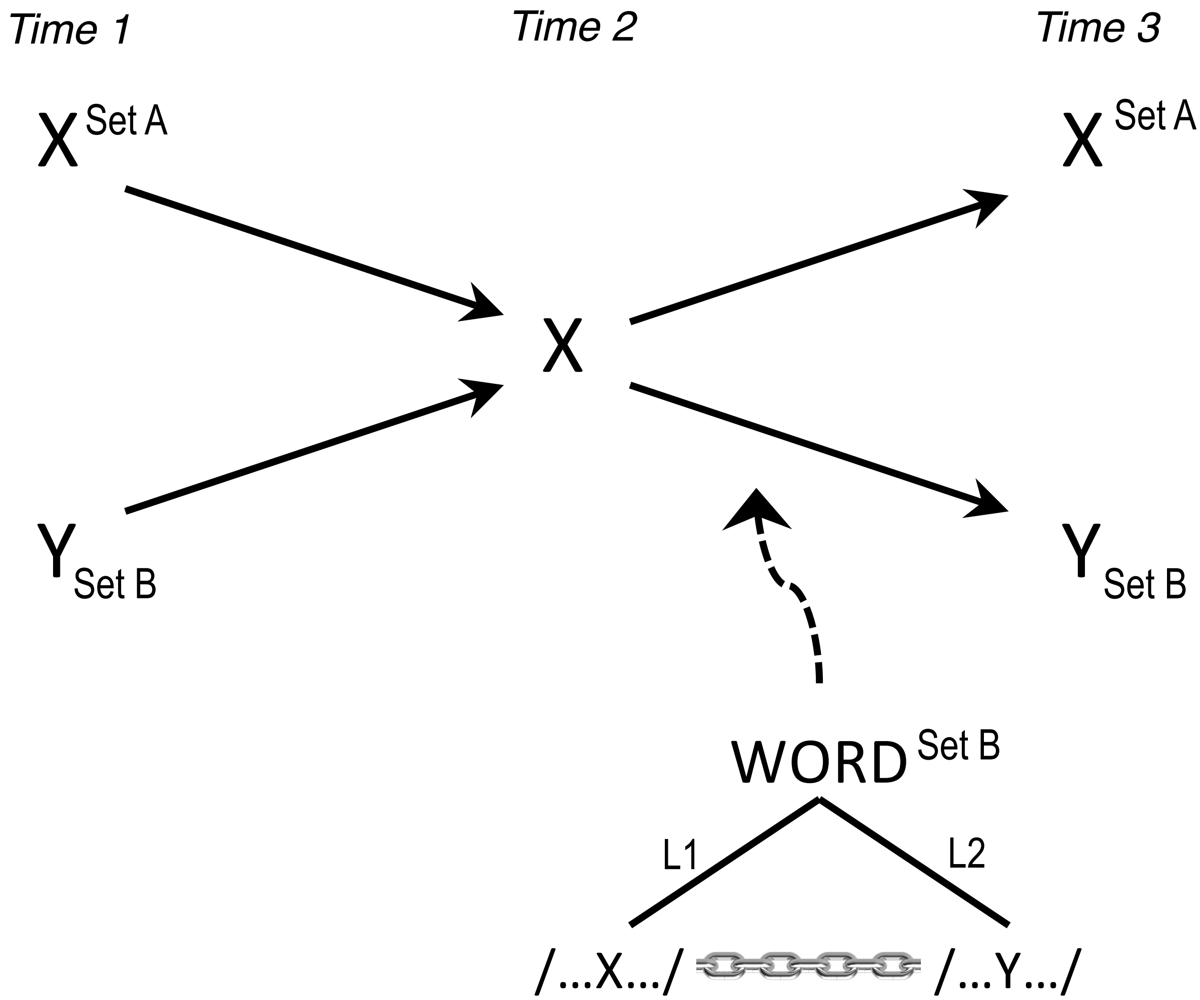

FIGURE 8. Illustration of the general mechanics of merger reversal in the L1 induced by crosslinguistic influence from an L2.

APPENDIX A. Materials and models in Study 1.

Item	**Lexical set**	**Shanghainese citation form* (in Stage II)**	**Mandarin form**	**Embedding compound (and part of speech)**	**Embedding compound frequency**
雷	MN-[ej]	le˨˧	lej˧˥	打雷 'thunder strikes' (v.)	High
垒	MN-[ej]	le˨˧	lej˨˩˦	堡垒 'fortress' (n.)	Low
缆	MN-[an]	lɛ˨˧	lan˨˩˦	光缆 'optical fiber' (n.)	High
澜	MN-[an]	lɛ˨˧	lan˧˥	狂澜 'huge wave' (n.)	Low
来	MN-[aj]	lɛ˨˧	laj˧˥	上来 'to come up' (v.)	High
睐	MN-[aj]	lɛ˨˧	laj˥˩	青睐 'to favor' (v.)	Low
配	MN-[ej]	p^he˧˦	p^hej˥˩	搭配 'to match with' (v.)	High
沛	MN-[ej]	p^he˥˧	p^hej˥˩	充沛 'abundant' (adj.)	Low
滩	MN-[an]	t^hɛ˥˧	t^han˥	外滩 'the Bund' (n.)	High
坍	MN-[an]	t^hɛ˥˧	t^han˥	压坍 'to crash' (v.)	Low
态	MN-[aj]	t^hɛ˧˦	t^haj˥˩	状态 'status' (n.)	High
胎	MN-[aj]	t^hɛ˥˧	t^haj˥	保胎 'to protect the fetus' (v.)	Low
贝	MN-[ej]	pe˧˦	pej˥˩	宝贝 'treasure' (n.)	High
狈	MN-[ej]	pe˧˦	pej˥˩	狼狈 'in an extremely embarrassing state' (adj.)	Low
班	MN-[an]	pɛ˥˧	pan˥	上班 'to go to work' (v.)	High
阪	MN-[an]	pɛ˧˦	pan˨˩˦	大阪 'Osaka (Japanese city)' (n.)	Low
呆	MN-[aj]	tɛ˥˧	taj˥	痴呆 'retarded' (adj.)	High
歹	MN-[aj]	tɛ˥˧	taj˨˩˦	为非作歹 'to do bad things' (v.)	Low

* The tone of the test items will change when embedded in a compound due to tone sandhi.

TABLE A1. Critical items in Study 1.

	LM model on F1Start				LM model on F2Start			
	β	**S.E.**	**t**	**p_{MCMC}**	**β**	**S.E.**	**t**	**p_{MCMC}**
(Intercept)	634.01	8.58	73.91	**<.001**	1530.92	13.05	117.33	**<.001**
Age=Old	-54.20	3.49	-15.55	**<.001**	26.13	3.07	8.52	**<.001**
Sex=F	37.90	2.58	14.70	**<.001**	91.72	2.35	39.12	**<.001**
Onset=L	46.90	8.01	5.86	**<.001**	-47.57	7.34	-6.48	**<.001**
Onset=PHTH	-13.86	8.01	-1.73	.097	44.53	7.34	6.07	**<.001**
	LM model on F1End				**LM model on F2End**			
	β	**S.E.**	**t**	**p_{MCMC}**	**β**	**S.E.**	**t**	**p_{MCMC}**
(Intercept)	634.72	14.88	43.26	**<.001**	1530.40	15.62	98.05	**<.001**
LexSet=MN-[an]	-5.65	15.24	-0.37	.72	21.87	15.89	1.38	.15
LexSet=MN-[ej]	-74.17	15.23	-4.87	**<.001**	66.74	15.88	4.20	**<.001**
Age=Old	-31.73	4.18	-7.59	**<.001**	29.09	5.58	5.21	**<.001**
LexSet=MN-[an]: Age=Old	--	--	--	--	-18.20	7.83	-2.32	**.020**
LexSet=MN-[ej]: Age=Old	--	--	--	--	-24.95	7.79	-3.21	**.002**
Sex=F	28.12	3.15	8.93	**<.001**	91.00	2.44	37.26	**<.001**
Onset=L	28.87	8.80	3.28	**.0042**	--	--	--	--
Onset=PHTH	-11.76	8.80	-1.34	.18	--	--	--	--

bold = $p_{MCMC} < .05$

TABLE A2. Fixed-effect terms in the LM models on formant measures in the reading experiment, Study 1.

	β	**S.E.**	*z*	***p*(|*z*|)**
(Intercept)	-1.63	0.25	-6.42	**<.001**
LexSet=MN-[an]	0.50	0.25	2.00	**.045**
LexSet=MN-[ej]	2.42	0.25	9.75	**<.001**
Age=Old	-0.52	0.14	-3.63	**<.001**
Frq=H	0.33	0.10	3.38	**<.001**
Onset=L	1.14	0.14	8.38	**<.001**
Onset=PHTH	-0.94	0.15	-6.35	**<.001**

bold = $p(|z|) < .05$

TABLE A3. Fixed-effect terms in the GLM model on Diphthong in the reading experiment, Study 1.

	LM model on F1Start				**LM model on F2Start**			
	β	**S.E.**	***t***	p_{MCMC}	***β***	**S.E.**	***t***	p_{MCMC}
(Intercept)	663.30	9.37	70.75	**<.001**	1550.92	13.11	118.28	**<.001**
Age=Old	-53.74	3.71	-14.50	**<.001**	--	--	--	--
Sex=F	41.50	2.78	14.94	**<.001**	90.87	2.18	41.70	**<.001**
Onset=L	38.96	7.52	5.18	**<.001**	-37.91	6.47	-5.86	**<.001**
Onset=PHTH	-1.50	7.52	-0.20	.84	35.19	6.47	5.44	**<.001**
Block.L	7.91	3.19	2.48	**.014**	-6.78	2.45	-2.76	**.0062**
Block.Q	-0.96	3.19	-0.30	.77	-1.01	2.45	-0.41	.70
	LM model on F1End				**LM model on F2End**			
	β	**S.E.**	***t***	p_{MCMC}	***β***	**S.E.**	***t***	p_{MCMC}
(Intercept)	675.13	18.33	36.84	**<.001**	1566.59	15.70	99.81	**<.001**
LexSet=MN-[an]	-5.66	19.73	-0.29	.77	1.39	11.38	0.12	.89
LexSet=MN-[ej]	-130.14	19.73	-6.60	**<.001**	86.36	11.38	7.79	**.001**
Age=Old	-37.73	7.92	-4.76	**<.001**	7.51	5.59	1.34	.18
LexSet=MN-[an]: Age=Old	9.34	11.14	-0.84	.40	-0.95	7.86	-0.12	.89
LexSet=MN-[ej]: Age=Old	34.99	11.14	3.14	**.002**	-43.24	7.86	-5.50	**<.001**
Sex=F	39.33	3.47	11.32	**<.001**	92.08	2.48	37.07	**<.001**

bold = $p_{MCMC} < .05$

TABLE A4. Fixed-effect terms in the LM models on formant measures in the translation experiment, Study 1.

	β	S.E.	*z*	*p*(|*z*|)
(Intercept)	-1.38	0.26	-5.37	**<.001**
LexSet=MN-[an]	0.35	0.24	1.44	.15
LexSet=MN-[ej]	3.12	0.27	11.60	**<.001**
Age=Old	0.10	0.24	0.42	.68
LexSet=MN-[an]: Age=Old	-0.53	0.34	-1.55	.12
LexSet=MN-[ej]: Age=Old	-0.91	0.34	-2.64	**.008**
Frq=H	0.17	0.07	2.35	**.019**
Onset=L	0.97	0.11	9.20	**<.001**
Onset=PHTH	-0.56	0.11	-5.22	**<.001**

bold = $p(|z|) < .05$

TABLE A5. Fixed-effect terms in the GLM model on Diphthong in the translation experiment, Study 1.

APPENDIX B. Materials and models in Study 2.

Item	Lexical set	Shanghainese citation form* (in Stage II)	Mandarin form	Embedding compound (and part of speech)	Embedding compound frequency
退	Structure-mismatched MN-[ej]	tʰe˦	tʰwej˥˩	辞退 'to lay off' (v.)	High
腿	Structure-mismatched MN-[ej]	tʰe˦	tʰwej˨˩˦	方腿 'spam (meat)' (n.)	High
对	Structure-mismatched MN-[ej]	te˦	twej˥˩	不对 'not correct' (adj.)	High
碎	Structure-mismatched MN-[ej]	se˦	swej˥˩	打碎 'to break something' (v.)	High
配	Regular MN-[ej]	pʰe˦	pʰej˥˩	搭配 'to match with' (v.)	High
贝	Regular MN-[ej]	pe˦	pej˥˩	宝贝 'treasure' (n.)	High
态	MN-[aj]	tʰɛ˦	tʰaj˥˩	状态 'status' (n.)	High
呆	MN-[aj]	tɛ˥˧	taj˥	痴呆 'retarded' (adj.)	High
赛	MN-[aj]	sɛ˦	saj˥˩	决赛 'final competition' (n.)	High

* The tone of the test items will change when embedded in a compound due to tone sandhi.

TABLE B1. Critical items in Study 2.

	LM model on F1Start				LM model on F2Start			
	β	**S.E.**	***t***	p_{MCMC}	***β***	**S.E.**	***t***	p_{MCMC}
(Intercept)	583.56	8.72	66.91	**<.001**	1582.71	19.57	80.88	**<.001**
LexSet=MN-[aj]	--	--	--	--	-21.87	21.65	-1.01	.39
LexSet=MN-[ej] (regular)	--	--	--	--	-31.14	24.54	-1.27	.19
Age=Old	-50.66	4.83	-10.50	**<.001**	13.82	4.02	3.44	**<.001**
Sex=F	36.37	3.49	10.44	**<.001**	82.82	3.82	21.71	**<.001**
Onset=PHTH	26.46	6.54	4.05	**.006**	--	--	--	--
Onset=PT	-2.17	6.95	-0.31	.76	--	--	--	--
LexSet=MN-[aj]: Sex=F	--	--	--	--	12.75	4.74	2.69	**.007**
LexSet=MN-[ej] (regular): Sex=F	--	--	--	--	11.64	5.33	2.18	**.029**
	LM model on F1End				**LM model on F2End**			
	β	**S.E.**	***t***	p_{MCMC}	***β***	**S.E.**	***t***	p_{MCMC}
(Intercept)	547.98	13.60	40.29	**<.001**	1594.90	20.73	76.93	**<.001**
LexSet=MN-[aj]	78.16	10.53	7.42	**<.001**	-52.35	25.90	-2.02	**.042**
LexSet=MN-[ej] (regular)	-31.82	12.02	-2.65	**.036**	37.35	29.36	1.27	.17
Age=Old	-20.29	9.11	-2.23	**.028**	-3.86	6.44	-0.60	.54
LexSet=MN-[aj]: Age=Old	-34.64	14.08	-2.46	**.015**	38.70	9.99	3.87	**<.001**
LexSet=MN-[ej] (regular): Age=Old	37.65	15.77	2.39	**.016**	-16.77	11.18	-1.50	.14
Sex=F	31.43	4.58	6.86	**<.001**	81.80	4.12	19.85	**<.001**
Onset=PHTH	17.01	4.67	3.64	**.021**	--	--	--	--
Onset=PT	-8.42	4.97	-1.69	.16	--	--	--	--
LexSet=MN-[aj]: Sex=F	--	--	--	--	12.60	5.16	2.44	**.015**
LexSet=MN-[ej] (regular): Sex=F	--	--	--	--	7.34	5.80	1.26	.22

bold = $p_{MCMC} < .05$

TABLE B2. Fixed-effect terms in the LM models on formant measures in the reading experiment, Study 2.

	β	**S.E.**	***z***	***p*(|*z*|)**
(Intercept)	-0.26	0.30	-0.87	0.38
LexSet=MN-[aj]	-2.53	0.44	-5.69	**<.001**
LexSet=MN-[ej] (regular)	1.10	0.38	2.92	**.004**
Age=Old	-0.94	0.27	-3.46	**<.001**
LexSet=MN-[aj]: Age=Old	1.46	0.57	2.57	**.010**
LexSet=MN-[ej] (regular): Age=Old	-0.71	0.47	-1.49	.14
Onset=PHTH	-0.14	0.16	-0.83	.41
Onset=PT	0.64	0.17	3.71	**<.001**

bold = $p(|z|) < .05$

TABLE B3. Fixed-effect terms in the GLM model on Diphthong in the reading experiment, Study 2.

	LM model on F1Start				**LM model on F2Start**			
	β	**S.E.**	***t***	p_{MCMC}	***β***	**S.E.**	***t***	p_{MCMC}
(Intercept)	622.65	8.44	73.81	**<.001**	1574.97	14.85	106.06	**<.001**
Age=Old	-56.69	5.69	-9.96	**<.001**	--	--	--	--
Sex=F	47.37	4.04	11.72	**<.001**	81.79	2.88	28.38	**<.001**
Onset=PHTH	43.28	5.32	8.14	**<.001**	--	--	--	--
Onset=PT	-8.60	5.67	-1.52	.17	--	--	--	--
	LM model on F1End				**LM model on F2End**			
	β	**S.E.**	***t***	p_{MCMC}	***β***	**S.E.**	***t***	p_{MCMC}
(Intercept)	539.38	16.43	32.82	**<.001**	1657.96	16.46	100.76	**<.001**
LexSet=MN-[aj]	112.90	10.70	10.55	**<.001**	-87.78	9.15	-9.60	**<.001**
LexSet=MN-[ej] (regular)	7.47	12.23	0.61	.57	5.82	10.35	0.56	0.59
Age=Old	-4.37	9.83	-0.44	.66	-33.61	6.01	-5.60	**<.001**
LexSet=MN-[aj]: Age=Old	-41.66	15.03	-2.77	**.007**	47.67	9.18	5.20	**<.001**
LexSet=MN-[ej] (regular): Age=Old	1.44	16.97	0.08	0.93	-18.47	10.37	-1.78	.079
Sex=F	42.98	4.97	8.66	**<.001**	92.30	3.09	29.88	**<.001**
Onset=PHTH	23.62	4.51	5.24	**.001**	--	--	--	--
Onset=PT	-9.64	4.87	-1.98	.10	--	--	--	--

bold = $p_{MCMC} < .05$

TABLE B4. Fixed-effect terms in the LM models on formant measures in the translation experiment, Study 2.

	β	**S.E.**	***z***	***p*(|*z*|)**
(Intercept)	1.02	0.37	2.74	**.006**
LexSet=MN-[aj]	-3.34	0.38	-8.76	**<.001**
LexSet=MN-[ej]	0.50	0.38	1.33	.18
Age=Old	-0.97	0.28	-3.50	**<.001**
LexSet=MN-[aj]: Age=Old	1.36	0.50	2.73	**.006**
LexSet=MN-[ej] (regular): Age=Old	-0.11	4.50	-0.23	.82

bold = $p(|z|) < .05$

TABLE B5. Fixed-effect terms in the GLM model on Diphthong in the translation experiment, Study 2.

APPENDIX C. Materials and models in Study 3.

Item	Lexical set	Shanghainese citation form* (in Stage II)	Mandarin form	Embedding compound (and part of speech)	Embedding compound frequency
赔	Onset-mismatched MN-[ej]	be˨˧	p^{h}ej˧˥	索赔 ‘to ask for indemnification’ (v.)	Low
陪	Onset-mismatched MN-[ej]	be˨˧	p^{h}ej˧˥	不陪 ‘not to accompany’ (v.)	Low
备	Onset-mismatched MN-[ej]	be˨˧	pej˥˩	准备 ‘to prepare’ (v.)	High
倍	Onset-mismatched MN-[ej]	be˨˧	pej˥˩	两倍 ‘twice’ (adj.)	High
配	Regular MN-[ej]	p^{h}e˧˦	p^{h}ej˥˩	搭配 ‘to match with’ (v.)	High
沛	Regular MN-[ej]	p^{h}e˥˨	p^{h}ej˥˩	充沛 ‘abundant’ (adj.)	Low
贝	Regular MN-[ej]	pe˧˦	pej˥˩	宝贝 ‘treasure’ (n.)	High
狈	Regular MN-[ej]	pe˧˦	pej˥˩	狼狈 ‘in an extremely embarrassing state’ (adj.)	Low
态	MN-[aj]	t^{h}ɛ˧˦	t^{h}aj˥˩	状态 ‘status’ (n.)	High
胎	MN-[aj]	t^{h}ɛ˥˨	t^{h}aj˥	保胎 ‘to protect the fetus’ (v.)	Low
呆	MN-[aj]	tɛ˥˨	taj˥	痴呆 ‘retarded’ (adj.)	High
歹	MN-[aj]	tɛ˥˨	taj˨˩˦	为非作歹 ‘to do bad things’ (v.)	Low

* The tone of the test items will change when embedded in a compound due to tone sandhi.

TABLE C1. Critical items in Study 3.

	LM model on F1Start				LM model on F2Start			
	β	**S.E.**	***t***	***p*MCMC**	***β***	**S.E.**	***t***	***p*MCMC**
(Intercept)	630.99	10.75	58.70	**<.001**	1512.92	21.08	71.76	**<.001**
LexSet=MN-[aj]	-21.17	11.10	-1.91	.11	50.17	19.80	2.53	**.021**
LexSet=MN-[ej] (regular)	-20.86	11.10	-1.88	.11	22.45	19.80	1.13	.26
Age=Old	-56.82	4.00	-14.19	**<.001**	23.93	4.01	5.97	**<.001**
Sex=F	36.15	2.93	12.33	**<.001**	82.42	4.20	19.62	**<.001**
Onset=PHTH	13.74	4.97	2.76	**.030**	17.66	8.84	2.00	.057
Frq=H	14.90	7.01	2.12	.075	--	--	--	--
LexSet=MN-[aj]: Frq=H	-18.25	9.95	-1.83	.11	--	--	--	--
LexSet=MN-[ej] (regular): Frq=H	-27.66	9.92	-2.79	**.028**	--	--	--	--
LexSet=MN-[aj]: Sex=F	--	--	--	--	12.08	5.04	2.40	**.018**
LexSet=MN-[ej] (regular): Sex=F	--	--	--	--	9.94	5.01	1.98	.053
	LM model on F1End				**LM model on F2End**			
	β	**S.E.**	***t***	***p*MCMC**	***β***	**S.E.**	***t***	***p*MCMC**
(Intercept)	583.97	16.77	34.82	**<.001**	1572.45	23.71	66.31	**<.001**
LexSet=MN-[aj]	37.12	17.03	2.18	.055	-33.39	29.72	-1.12	.20
LexSet=MN-[ej] (regular)	-18.35	17.01	-1.08	.31	14.88	29.71	0.50	.56
Age=Old	-24.13	5.52	-4.37	**<.001**	-0.43	7.01	-0.06	.97
Sex=F	27.22	4.14	6.58	**<.001**	88.05	3.14	28.03	**<.001**
LexSet=MN-[aj]: Age=Old	--	--	--	--	31.57	10.09	3.13	**.002**
LexSet=MN-[ej] (regular): Age=Old	--	--	--	--	12.33	9.99	1.23	.22

bold = p_{MCMC} < .05

TABLE C2. Fixed-effect terms in the LM models on formant measures in the reading experiment, Study 3.

	β	**S.E.**	***z***	***p*(|*z*|)**
(Intercept)	-0.44	0.45	-0.96	.33
LexSet=MN-[aj]	-2.23	0.50	-4.49	**<.001**
LexSet=MN-[ej] (regular)	0.85	0.47	1.82	.069
Age=Old	-0.79	0.18	-4.45	**<.001**
Sex=F	0.29	0.13	2.21	**.027**
Onset=PHTH	-0.45	0.23	-2.00	**.046**

bold = $p(|z|) < .05$

TABLE C3. Fixed-effect terms in the GLM model on Diphthong in the reading experiment, Study 3.

	LM model on F1Start				LM model on F2Start			
	β	**S.E.**	***t***	p_{MCMC}	***β***	**S.E.**	***t***	p_{MCMC}
(Intercept)	665.84	8.22	81.00	**<.001**	1538.19	18.06	85.20	**<.001**
LexSet=MN-[aj]	-32.66	6.50	-5.03	**.002**	44.37	15.41	2.88	**.004**
LexSet=MN-[ej] (regular)	-8.00	6.47	-1.24	.29	12.52	15.40	0.81	.42
Age=Old	-60.63	4.76	-12.73	**<.001**	--	--	--	--
Sex=F	46.89	3.41	13.75	**<.001**	83.42	2.69	31.01	**<.001**
Onset=PHTH	22.56	2.89	7.79	**<.001**	13.45	6.89	1.95	.051
Frq=H	-0.57	4.10	-0.14	.91	--	--	--	--
LexSet=MN-[aj]: Frq=H	4.04	5.81	0.69	.54	--	--	--	--
LexSet=MN-[ej] (regular): Frq=H	-11.98	5.78	-2.07	.077	--	--	--	--
	LM model on F1End				**LM model on F2End**			
	β	**S.E.**	***t***	p_{MCMC}	***β***	**S.E.**	***t***	p_{MCMC}
(Intercept)	574.42	17.45	32.91	**<.001**	1648.40	16.46	100.15	**<.001**
LexSet=MN-[aj]	77.97	11.79	6.61	**<.001**	-69.39	10.77	-6.44	**<.001**
LexSet=MN-[ej] (regular)	-8.96	11.76	-0.76	.45	-3.11	10.75	-0.29	.79
Age=Old	2.44	9.83	0.25	.82	-40.42	7.04	-5.74	**<.001**
Sex=F	39.10	4.31	9.07	**<.001**	92.10	3.12	29.56	**<.001**
Onset=PHTH	15.80	4.28	3.69	**.005**	--	--	--	--
LexSet=MN-[aj]: Age=Old	-44.27	13.87	-3.19	**.001**	48.08	9.93	4.84	**<.001**
LexSet=MN-[ej] (regular): Age=Old	-22.81	13.81	-1.65	.10	8.86	9.89	0.90	.38

bold = $p_{MCMC} < .05$

TABLE C4. Fixed-effect terms in the LM models on formant measures in the translation experiment, Study 3.

	β	**S.E.**	***z***	***p*(|*z*|)**
(Intercept)	1.33	0.38	3.53	**<.001**
LexSet=MN-[aj]	-3.61	0.35	-10.26	**<.001**
LexSet=MN-[ej] (regular)	0.14	0.31	0.46	.64
Age=Old	-1.56	0.29	-5.40	**<.001**
LexSet=MN-[aj]: Age=Old	1.99	0.46	4.36	**<.001**
LexSet=MN-[ej] (regular): Age=Old	0.68	0.41	1.66	.096

bold = $p(|z|) < .05$

TABLE C5. Fixed-effect terms in the GLM model on Diphthong in the translation experiment, Study 3.