

## Long-lag identity priming in the absence of long-lag morphological priming: evidence from Mandarin tone alternation

**Stephen Politzer-Ahles**, University of Kansas, United States; The Hong Kong Polytechnic University, Hong Kong, [sjpa@ku.edu](mailto:sjpa@ku.edu)

**Lei Pan**, The Hong Kong Polytechnic University, Hong Kong; Nanyang Technological University, Singapore, [PANL0006@e.ntu.edu.sg](mailto:PANL0006@e.ntu.edu.sg)

**Jueyao Lin**, The Hong Kong Polytechnic University, Hong Kong, [jueyao-hansy.lin@connect.polyu.hk](mailto:jueyao-hansy.lin@connect.polyu.hk)

**Ka Keung Lee**, The Hong Kong Polytechnic University, Hong Kong, [leon-leechk@hotmail.com](mailto:leon-leechk@hotmail.com)

The present study tested whether listeners hearing one form of a morpheme activate other forms of the same morpheme. Listeners performed lexical decisions while hearing Mandarin monosyllables; crucially, critical targets could be primed by related syllables that occurred 18–52 trials earlier (long-lag priming). The use of long-lag priming ensures that any facilitation effects are due to morphological relatedness and not to semantic or form relationships, which do not prime lexical decisions at long lags. Across three experiments (total N = 458), we consistently found that lexical decisions were primed when the same pronunciation of a morpheme occurred as prime and target (e.g.,  $shi^l - shi^l$ ) but were not primed when two different variants of the same morpheme occurred as prime and target (e.g.,  $shi^R - shi^l$ , where both of these syllables are potential pronunciations of the same morpheme). In other words, we observed identity priming but not morphological priming, unlike other long-lag priming experiments, which almost invariably observe intramodal morphological priming if they test it. This surprising finding suggests that there are boundary conditions on the elicitation of long-lag morphological priming effects.



## 1. Introduction

In order to use and understand language, people need to use physical input (such as speech sounds, written forms, or signs) to access linguistic representations (such as words, their meanings, and their grammatical properties). How this process of word recognition or lexical activation works has been the focus of a large amount of research for the past half century. However, a complication to this issue is the fact that a linguistic expression, such as a word or a morpheme, can be pronounced in many different ways. These differences can be due to idiosyncratic factors like different people's voices, emotional state, and random variation. They can also be due to systematic linguistic factors, such as context-conditioned phonological changes – for example, the plural morpheme *-s* in English is pronounced as [ɪz] after a sibilant (e.g., *buses*), as [z] after a non-sibilant voiced sound (e.g., *dogs* or *zebras*), and as [s] after a non-sibilant voiceless sound (e.g., *cats*). Because of variations like these, a language user needs to be able to map many different inputs onto the same lexical representation – if someone cannot recognize that, e.g., *cat* pronounced in two different voices is still the same word, then they would not be able to comprehend language, because every stimulus they hear would appear to be a new and unknown word. In the present paper we are most interested in examining how systematic phonological factors, like the abovementioned alternation pattern, are used to recognize words. To preview: we will find that there are circumstances in which word recognition does not seem to take systematic phonological alternations into account.

One of the main methods for studying how words are represented in the mind, and how physical stimuli make contact with these mental representations, is *priming*. Encountering some linguistic stimulus once usually helps people process the same stimulus more easily when they encounter it again – i.e., the first encounter with the stimulus *primes* the next one. For example, if a person needs to respond to a word (referred to as a “target”) as quickly as possible when they hear it, they can respond more quickly if they have recently heard the same word than if they have not – e.g., people can respond faster to *nurse* after having heard *nurse* than after having heard some other word. Crucially, this priming effect does not always rely on encountering the exact same word twice, but can also occur when a person had encountered a similar word – in other words, not only does *nurse* prime *nurse*, but, under certain circumstances, other related words like *doctor*, *nursing*, or *purse* might also prime *nurse*. Thus, by revealing what types of primes facilitate what types of targets under what conditions, it is possible to infer what sorts of representations mediate the processing of primes and targets.

As mentioned above, language users must be able to recognize words and morphemes in spite of variability in the input signal; thus, if priming reflects the activation of lexical entries or at least their forms, then exposure to one word or morpheme should facilitate later processing of the same word or morpheme even if it occurs in a different form. Priming research has indeed demonstrated this; for example, hearing *bed* pronounced in one voice facilitates responses to *bed*

pronounced in another voice (Luce & Lyons, 1998), and hearing *assumption* facilitates responses to *assumed* even though the morpheme that these two words share is spelled and pronounced differently (Downie et al., 1985; Fowler et al., 1985; Napps, 1989). This pattern has also been observed with optional allophonic variation. For example, McLennan and colleagues (2003) found that exposure to *raider* pronounced with a casually articulated flap [ɾ] facilitated later responses to the same word carefully articulated with a [d], and vice versa, at least under certain experimental conditions.

There are several different views on how this mapping between different variants occurs. One possibility is that lexical representations are abstract and that a person uses linguistic knowledge to abstract away from the surface details of the input stimuli, and/or to undo phonological changes, in order to match different stimuli onto the same abstract representation: for example, a listener hears “raider” pronounced with a [ɾ], recognizes that English phonology allows a /d/ or /t/ to be flapped in certain positions including the position where the flap appears in this word, and thus infers that the flap could have been a realization of /d/, and from there infers that the input means *raider*. Models of lexical representation that adopt these sorts of assumptions are called “abstractionist” models. Another possibility is that lexical representations are episodic, and the “representation” of a given word is not a single form, but a collection of many instances of that word a person has heard or seen throughout their lifetime, such that variation in the surface forms of a word is directly coded in the lexicon; for example, a listener has heard *raider* pronounced with a flap enough times, and has stored those examples in memory, so that the next time they hear the flapped form they can recognize these and remember that they are a possible way of pronouncing *raider*. Models like this are called “episodic” or “instance-based” models. These two classes of models are not mutually exclusive (Wheeldon & Monsell, 1992; Luce et al., 2003), and modern accounts often assume that the lexicon needs to include a combination of abstract representations and information about frequency of variants and morphological components (e.g. Hay & Baayen, 2005; Connine & Pinnow, 2006; Sciana & Dowker, 2007; Pierrehumbert, 2016; Zuraw et al., 2021). Even if normal lexical access does involve episodic representations, some sort of abstraction presumably needs to also be available, because compensation for phonological alternation occurs even in pseudowords – for example, in French, where obstruents can assimilate in voicing to a following obstruent, listeners not only interpret e.g. [apsyrd] as the real word /absyrd/ *absurde* (Hallé et al., 2000), but also interpret e.g. [ofbe] as the pseudoword /ovbe/ (Sun et al., 2015; see also Tavabi et al., 2009). Importantly, both classes of models predict that different physical inputs can make contact with the same lexical representations; they just differ in terms of what the nature of those representations is (for more detailed reviews see, e.g., Tenpenny, 1995; Bowers, 2000; Bowers & Kouider, 2003; Connine & Pinnow, 2006; Kouider & Dupoux, 2009; Burt & Long, 2011; Pierrehumbert, 2016). Our goal in the present study is not to adjudicate between these accounts; here we are just highlighting

that there are multiple possible explanations for how priming between different phonological variants may occur.

The present study examines priming between different tonal variants of Mandarin syllables (described in more detail in 1.2 below). This manipulation differs in several ways from the abovementioned studies. Most obviously, tone is suprasegmental, whereas almost all previous priming research on phonological variation has focused on segmental alternations. Given that tone contributes to word recognition in a different way than segments do (e.g. Wiener & Turnbull, 2015), it is valuable to examine whether priming occurs across tonal variants in the same way it does across segmental variants. The main contribution of the present study is to examine activation of tonal variants using long-lag priming, which, for reasons that will be described below, is arguably a more appropriate paradigm than the immediate priming paradigm which previous studies of tonal alternation have used.

## 1.1 Priming of Mandarin tonal variants

Most morphemes in Mandarin are one syllable long and have either a High, Rising, Low, or Falling tone. Syllables with Low tone (also called third tone), like *qi<sup>L</sup>* “start”, are instead pronounced with something similar to a Rising tone (also called second tone) if they are followed by another Low tone within the same intonational unit, as in *qi<sup>R</sup> dian<sup>L</sup>* “start-point” (“starting point”);<sup>1</sup> this change is known as “third tone sandhi” (see, e.g., Zhang, 2010, among others). In other words, any Mandarin morpheme with underlying Low tone has multiple possible pronunciations – multiple allomorphs – depending on the phonological context, just like we saw above for English plural *-s*. Likewise, a surface Rising tone might be a realization of either an underlying Low tone or an underlying Rising tone.

The crucial question for us is whether hearing one Low- or Rising-tone syllable in isolation causes listeners to activate both Low and Rising underlying forms. One could imagine at least two possibilities. First, there might be no reason for a listener to activate multiple forms, because the input already unambiguously matches some canonical form in the lexicon,<sup>2</sup> and because there is no context to make the listener suspect that this surface form might have been changed from some other underlying form. On the other hand, listeners might activate multiple representations that the input corresponds to anyway; for example, if different allomorphs of each syllable are

---

<sup>1</sup> Transcriptions of segmental syllables given here are in Hanyu Pinyin; the superscript letters represent tone categories (H = High, R = Rising, L = Low, F = Falling).

<sup>2</sup> This is different from situations such as the English flapping example above, where [r] is not the canonical form of any English phoneme, and thus hearing a word with a flap may obligate the listener to recognize it as a variant of /t/ or /d/ in order to associate it to some canonical form. While not all theories of lexical access would agree that the phonologically “underlying” form is privileged in processing, see Sumner and Samuel (2005) and Ranbom and Connine (2007) for tentative evidence that canonical forms may indeed be processed differently than others.

listed in the lexicon (as has been proposed by, e.g., Nixon et al., 2014, and Li & Chen, 2015) or if the lexicon itself is made up of detailed memories of many episodes in which each item was experienced (Tenpenny, 1995; Goldinger, 1998) and lexical entries are sound-meaning pairings reinforced by experience (Pulvermüller, 1999), then the recognition of one form of a morpheme may also cause activation of related forms of the same morpheme. If so, then a listener hearing a Rising tone might also activate Low-tone variants of the same syllable (because they have experience hearing syllables with the same meaning in that tone, or because they recognize that this Rising-tone input is sometimes a realization of the Low-tone form), and likewise a listener hearing a Low tone might also activate Rising-tone variants of the same syllable (because, again, they have experience hearing the same morpheme in both tones, or because they recognize that this Low-tone form can sometimes be realized with Rising tone).

Several recent studies have examined priming between Mandarin Low and Rising tones. Chien and colleagues (2016) examined the representation of Mandarin words including phonologically altered tones, like *qi<sup>R</sup> dian<sup>L</sup>* (underlyingly *qi<sup>L</sup> dian<sup>L</sup>*). They found that recognition of these words was primed when they were immediately preceded by presentation of the first morpheme in its underlying form (*qi<sup>L</sup>*) but were not primed by presentation of its surface form (*qi<sup>R</sup>*). This suggests that primes in isolation only activated phonologically matching lexical forms (i.e., Low tone only activated Low tone, and Rising tone only activated Rising tone), but then the presentation of the target which involved a surface Rising tone in a sandhi context involved recovery of the underlying Low tone. Thus, the Low tone recovered from the target was primed by the Low prime, and not by the Rising prime. It is, however, not clear that the recovery of the Low form from the surface Rising tone in the target is due to phonological knowledge. It is likely due instead to semantic or lexical knowledge, since the syllable with a surface Rising tone in the target is embedded within a compound word whose meaning is also related to the Low-tone prime and/or to many other words from the same cohort as the target (e.g., the target *xuan<sup>R</sup> ju<sup>L</sup>* ‘election’ is closely related to the prime *xuan<sup>L</sup>* ‘choose’, the target *fu<sup>R</sup> dao<sup>L</sup>* ‘to coach/tutor’ is closely related to other words sharing the same initial morpheme such as *fu<sup>L</sup> zhu<sup>F</sup>* ‘to assist’, etc.), not to mention that the initial syllable of the target is written with a character whose canonical pronunciation is Low-tone. Therefore, the priming observed here may be mainly due to the activation of the meaning of a known compound word, rather than purely due to activation of phonological variants.

Meng and colleagues (2021, experiment 3a) used a similar manipulation as Chien and colleagues (2016) but with a cross-modal rather than purely auditory paradigm, and with a shorter interval between prime and target. They found that visually-presented two-character words in which the first syllable would undergo sandhi were primed by both Low and Rising primes. Crucially, Rising primes only facilitated targets that began with underlying Low tones when those tones were in contexts that would trigger them to be produced as Rising (e.g., *kao<sup>R</sup>*

primed *kao<sup>R</sup> gu<sup>L</sup>* “archaeology” – which is underlyingly *kao<sup>L</sup> gu<sup>L</sup>* – but not *kao<sup>L</sup> cha<sup>R</sup>* “investigate” [experiment 1]), whereas Low primes facilitated both types of targets. This shows that the priming of sandhi-undergoing words by both Low and Rising targets is not just a result of Low and Rising being processed the same way, but is a result of the relationship between these tones and tone sandhi. In short, for Chien et al. (2016) only Low primed sandhi targets, whereas for Meng et al. (2021) both Low and Rising did. The difference between these findings is likely due to how the stimuli were presented in these experiments; in Meng and colleagues’ study the participants could begin processing the target while the prime was likely still being processed (see also Nixon et al., 2014, on the role of stimulus-onset asynchrony in priming Mandarin tone sandhi forms). In any case, the facilitation of sandhi-undergoing targets by Low-tone primes in this study is subject to the same caveat as described above for Chien and colleagues (2016), and the facilitation by Rising-tone primes is not necessarily evidence for activation across different variants since both the prime and target in this condition share the same surface form.

A recent study (Politzer-Ahles et al., 2022) failed to observe the sort of between-allomorph facilitation seen in the other studies described above. In an immediate priming paradigm, Mandarin-speaking participants heard a target syllable shortly after seeing either a Chinese character whose canonical pronunciation matches it or one whose pronunciation mismatches it in some way. Syllables whose tones mismatched the canonical pronunciation of the preceding character triggered some recognition difficulty (as evidenced by more negative event-related brain potentials) relative to syllables whose tones matched the preceding characters. Crucially, this effect was not modulated by the presence of a systematic phonological relationship between the heard target and the preceding character. Hearing a Rising tone when a Low tone was expected should not have caused much recognition difficulty because a Rising-tone form is actually a possible realization of an underlyingly Low-tone morpheme; however, this situation yielded just as much of a mismatch effect as hearing a Rising tone when some other wholly unrelated tone was expected. This happened regardless of whether the critical syllable was in a context that licensed the tone change. In other words, hearing an allomorph of the expected syllable was no different than hearing a phonologically unrelated syllable. This finding seems inconsistent with the others described above, as it suggests that hearing one form of a morpheme does not necessarily activate other phonological variants of this morpheme, at least when sentential context and most lexical context is removed. In that study, however, participants’ task was to judge whether the target matched the prime, which might not tap into the same mechanisms as normal word recognition (as opposed to the other studies reviewed above, which used a lexical decision task).

So, is it the case that exposure to one form of a Mandarin morpheme activates representations of its phonologically-conditioned allomorphs? While several previous results suggest that this can happen, some may not be due to the activation of phonologically-conditioned allomorphs

per se, but to other factors; furthermore, one set of recent results suggests that exposure to one form of a morpheme does not always activate its allomorphs, at least not when this allomorphy is based on a suprasegmental phonological change.

Because the abovementioned studies have all used a paradigm in which the target immediately follows the prime, it is difficult to disentangle contributions of form and meaning from contributions of morphology, as well as the role of response strategies (which might not necessarily reflect natural processing) on the part of participants. In the present study we use *long-lag priming*, which is arguably a more suitable paradigm for examining activation of morphemes while ruling out effects of form and meaning. The goal of the study was to use long-lag priming to examine whether facilitation occurs across different phonological forms of the same morpheme – i.e., whether hearing one form of a morpheme activates its other allomorphs. Below we review the relevant properties of long-lag priming, and then introduce the details of the present study.

## 1.2 Overview of long-lag priming

In long-lag priming (also known as long-term priming or delayed priming), other trials intervene between a participant's exposure to a prime and their response to a target. In other words, while a traditional paired priming study will have participants hear, e.g., *hunter* and then *hunt* and immediately respond to *hunt*, a long-lag priming study will have a participant hear some long series of words (and sometimes pseudowords) such as *hunter, washing, blosh, table, thurl, drem, hunt...*, while responding to each word. In some (particularly older) long-lag priming studies, participants explicitly “study” a prime, and later perform a task (typically some kind of metalinguistic judgment or verbal naming response) on the target. In others, participants perform the same task on both the prime and on the target, with extra trials intervening between them, as in the example above (see, e.g., Coane & Balota, 2010, and Horner & Henson, 2011, for discussion of how the tasks performed on primes and on targets may influence the priming effect). Under these conditions, participants are generally not aware that the experiment involves pairs of related stimuli – unlike in overt paired priming, where the presence of related vs. unrelated pairs is usually quite obvious and where participants may adjust their processing strategy accordingly. Furthermore, while targets that share a morpheme with (or are identical to) an earlier prime are facilitated in this paradigm, targets that share only a meaning or form relationship with an earlier prime are usually not facilitated (see, e.g., Feldman, 2000, for review); this makes the paradigm ideal for examining how different allomorphs can activate underlying morphemes while removing potential confounds due to semantic relationships or superficial form relationships.

Another paradigm with similar benefits as long-lag priming is subliminal masked priming, in which a prime immediately precedes a target but is presented for a short duration sandwiched between masks that make it unlikely for participants to be consciously aware of the prime. It also reduces the role of strategic factors, as participants are not aware of prime-target relationships



(since they're not aware of the primes at all), and it also preserves morphological and identity priming while rarely showing meaning or form priming. (In spite of these similarities, it should not be assumed that masked priming and long-lag priming measure the same process; see, e.g., Bowers, 2000, for arguments that they reflect different mechanisms.) However, it was not ideal for the present study because the present study focuses on variants of Mandarin morphemes, which necessitates presentation of spoken stimuli (as phonological alternations in Mandarin are not represented in the standard written form). While subliminal masked priming with spoken stimuli is possible (e.g., Kouider & Dupoux, 2005; Ussishkin et al., 2015), it requires substantially compressing the duration of the primes. Since duration is an important secondary cue for tone identity in Mandarin, compression may influence participants' recognition of the primes, which could substantially complicate the present study and the interpretation of its results. Using long-lag priming avoids this complication.

To set the stage for the manipulation that will be used in the present study, below we summarize some key properties of long-lag priming that are relevant for comparing phonologically alternated forms of Mandarin syllables.

### 1.2.1 Lack of long-lag semantic priming

Studies have consistently failed to observe long-lag facilitation of lexical decisions when primes have only a semantic or associative relationship to the target (Dannenbring & Briand, 1982; Feldman, 1992, 2000; Roediger & Challis, 1992; Zwitserlood et al., 2000, also find this pattern in picture naming). The only exception we are aware of is a study by Davelaar and Coltheart (1975), who found semantic priming when primes and targets were only separated by one intervening item; at longer lags, semantic priming is consistently absent, or present but non-significant (e.g., Becker et al., 1997, observed a non-significant 10-millisecond priming effect).

Relatedly, multiple studies have found that semantics does not contribute to morphological priming in long-lag paradigms; i.e., targets preceded by a prime which is related both semantically and morphologically to the target are not facilitated any more than targets preceded by a prime which is related only morphologically and not semantically (Bentin & Feldman, 1990; Feldman, 1992; Dohmes et al., 2004; Lensink et al., 2014). Feldman and Siok (1999) report a similar finding but with radicals (somewhat meaningful components of written Chinese characters) rather than morphemes. Priming has also been found between primes and targets that share an affix but whose whole-word meanings have little relationship (VanWagenen, 2014; Gaston et al., 2021).

Contra these results, however, Rueckl and Aicher (2008) found long-lag priming only for truly morphological relationships (such as *teacher-teach*) and not for apparent morphological relationships (such as *corner-corn*, which they treat as a semantically opaque morphological relationship); this could be argued to be a case of semantics contributing to long-lag priming.



Creemers and colleagues (2020) also found an effect in this direction (larger priming from primes related both semantically and morphologically, and from those that are only related semantically, than from primes that are only related morphologically), but the pattern was not significant and the lag between primes and targets was only five trials. Zhou and Marslen-Wilson (1995) and Tsang and colleagues (2014) found similar patterns (although this pattern is statistically non-significant in Tsang et al., 2014). In their studies, long-lag morphological priming for lexical decisions only obtained when the apparent morpheme shared between prime and target had the same meaning in both prime and target (e.g., 公園 ‘public park’ – 公廁 ‘public toilet’, where the character 公 in each word means ‘public’) and not when the apparent morpheme had different meanings in the prime and target (e.g., 公雞 ‘rooster’ [literally ‘male chicken’] – 公廁 ‘public toilet’, where the character 公 means ‘male’ in the prime but means ‘public’ in the target). These results also could be taken as evidence that semantics can contribute to long-lag priming.

### 1.2.2 Lack of long-lag form priming

Lexical decisions are generally not facilitated when primes at long lags have only a form (phonetic or orthographic) relationship with the target (Murrell & Morton, 1974; Napps & Fowler, 1987; Drews & Zwitserlood, 1995; Feldman & Moskovljević, 1987; Feldman, 2000; Gaston et al., 2021; this pattern also holds in picture naming, see Koester & Schiller, 2008). As with the lack of semantic priming, however, some studies do observe numerical priming effects that just fail to reach significance (Hanson & Wilkenfeld, 1985; Stoltz & Feldman, 1995; Zwitserlood et al., 2000). In written Mandarin, Zhou & Marslen-Wilson (1995) find no long-lag form priming in two-character words when the prime-target overlap is at the beginning of the target (i.e., the first character), but do find form priming when the overlap is at the end of the target; they argue that overlap at the beginning causes cohort competition but overlap at the end does not.

Form priming may have subtle or unacknowledged effects in long-lag priming, however, in that priming effects attributed to morphology may also be influenced by form. For example, Emmorey (1991) tested long-lag morphological priming in American Sign Language by priming base forms of verbs with forms of the same verb under different aspect or agreement inflection. For native signers, three of the four inflectional relationships in the experiment yielded significant priming; the one relationship that did not yield priming also happened to be the one in which the inflected form has perhaps the biggest perceptual difference from the base form. Thus, the apparent morphological priming in this study may be supported by a form relationship, or at least having a close form relationship may be a necessary precondition for observing morphological priming. (Form-related nonwords were also included in the study, but they yielded positive but non-significant priming, so they do not definitively rule out the possibility that form priming was occurring in the experiment.)

### 1.2.3 Presence of long-lag morphological priming

Studies have consistently found facilitation of lexical decision times at long lags when prime and target share a morpheme, whether the relationship is one of affixation or ablaut (Feldman & Fowler, 1987; Feldman & Moskovljević, 1987; Stoltz & Feldman, 1995; Marslen-Wilson & Tyler, 1998; Feldman, 2000; VanWagenen, 2014; Wilder et al., 2019; Gaston et al., 2021;), non-concatenative inflectional morphology (Emmorey, 1991), or compounding (Zhou & Marslen-Wilson, 1995); facilitation has also been observed on picture naming latencies (Zwitserslood et al., 2000; Koestler & Schiller, 2008; Lensink et al., 2014). Two rare exceptions are a study by Münte and colleagues (1999), in which morphological relationships yielded a priming effect in event-related brain potentials (see also Weyerts et al., 1996) but not in lexical decision reaction times, and one by De Grauwe and colleagues (2014), in which there was morphological priming in behavioral accuracy and brain activation for second-language learners of Dutch but not for native speakers.

As discussed above, morphological priming tends to hold even when the shared morpheme is spelled or pronounced differently in the prime and the target (e.g., Fowler et al., 1985; Boyce et al., 1987; Emmorey, 1991). Some studies that find morphological priming for regular inflection fail to find it for irregular morphological relationships where the morpheme is pronounced differently between prime and target (Kempley & Morton, 1982; Weyerts et al., 1996; Münte et al., 1999), but none of these studies used lexical decision times as the dependent measure.<sup>3</sup> Nevertheless, it is possible that morphological priming does require some minimum amount of form overlap. To our knowledge, no study has tested long-lag morphological priming with completely suppletive pairs (e.g., *go-went*) or highly dissimilar pairs (e.g., *teach-taught*), in contrast to the immediate priming literature, where these sorts of relationships have been a major focus (e.g., Stockall & Marantz, 2006). In long-lag priming studies, even the irregular relationships that have been tested are ones that still involve large form overlap (e.g., *assumed-assumption* and *clearly-clarify* in Fowler et al., 1985; *give-gave* in Marslen-Wilson & Tyler, 1998; etc.); thus, it remains to be seen whether long-lag morphological priming is possible between variants of a morpheme that share no form relationship at all. However, long-lag priming has been argued to be driven not by activation of lexical entries or lexemes themselves but by activation of their phonological or orthographic codes which occurs as language users attempt to map input stimuli onto meaningful mental representations (Bowers, 2000; Bowers & Kouider, 2003) – evidence for this view includes the facts that long-lag morphological priming is modality-dependent (i.e. auditory primes generally don't prime visual targets or vice versa; Bowers, 1996; Bowers & Michita, 1998) and occurs even for pseudowords. Under a view like this, it might be expected

---

<sup>3</sup> A lexical decision study which might have failed to find priming for irregular morphology is Stanners et al. (1979), but that study did not include an unrelated control condition, so all that can be concluded from it is that irregular relationships produced less priming than regular or identity relationships, not that they produced no priming at all.

that long-lag priming depends on having at least some overlap between the phonological or orthographic codes of the prime and the target.

Another important feature of long-lag morphological priming is that the facilitation from morphologically-related primes is often smaller than that for completely identical primes (Murrell & Morton, 1974; Emmorey, 1991; Feldman, 1992; Drews & Zwitserlood, 1995; Stoltz & Feldman, 1995); even when this pattern is not significant in individual studies, it is consistently in this same direction numerically (e.g., in Hanson & Wilkenfeld, 1985; for reviews see Tenpenny, 1995; Bowers, 2000; and Wilder et al., 2019).

#### 1.2.4 Chinese radicals

Feldman and Siok (1999) examined long-lag priming for Chinese radicals, which are somewhat meaningful sub-components of many Chinese characters. For example, a character like 打, which corresponds to a morpheme pronounced as *da*<sup>1</sup> and meaning ‘hit’ in isolation, is made up of two components. The right-hand part, 丁, is a phonetic component (also called a phonetic radical) which gives some hint about the character’s pronunciation (most characters with this component are pronounced with /t/ or /d/). The left-hand part, 扌, is a semantic component, also called a radical or a semantic radical; it represents a hand, and many characters with this component have something to do with actions that can be done with the hands (e.g., 推 ‘push’ and 拉 ‘pull’). While radicals like these are not morphemes, and differ from morphemes in several important ways (e.g., they are not accessible to syntax in the way that morphemes are, and thus are not “units” in any morphosyntactic sense), they also are like morphemes in some ways (e.g., both are meaningful, although the relationships between their meanings and the words they make up are sometimes quite opaque). Feldman and Siok (1999) found that lexical decision at long lags was facilitated by primes which share a radical with the target *and* have a related meaning. For primes that share a radical with the target but do not have a related meaning, they found a non-significant 10-millisecond priming effect; assuming a symmetrical confidence interval around this effect, that means their result is consistent with no long-lag priming effect for radicals, but also consistent with a priming effect as big as 20 milliseconds. So, even if one assumes that radicals are kind of like morphemes, this finding is unfortunately not very informative with respect to the question of whether morphemes (or radicals) alone contribute to long-lag priming (compare to the studies in other languages reviewed above that found morphological long-lag priming usually persists even without any semantic relationship).

#### 1.2.5 Translations

Translations are another interesting case, as cross-language translation equivalents arguably share a lexical representation (in bilinguals who know both languages) in spite of sometimes having very different forms. Exposure to a word in one language does not prime its translation

equivalents in other languages, however, unless the task performed in response to the prime forces the participant to think of the form of its translation equivalents (Kirsner et al., 1984; Durgunoğlu & Roediger, 1987) – for example, performing lexical decision to French words in one block does not prime lexical decisions to corresponding English words in a later block, but mentally translating French words into English in one block does prime later lexical decisions to those English words. If we assume that translation equivalents do share a lexical entry, these results suggest that long-lag morphological priming is mediated by form representations, rather than just reflecting repeated activation of the same lexical entry (see also the discussion in 1.2.3).

### 1.2.6 Episodic priming and the distinction between morphological and identity priming

As mentioned above, while morphological long-lag priming is quite reliable, it also tends to be smaller than repetition priming. This is surprising if we assume that both repetition and morphological priming (e.g., priming for *assume-assume* and priming for *assumption-assume*) each involve activation of phonological or orthographic codes that contact the same lexical entry.

One possible explanation for this discrepancy (articulated by, e.g., Tenpenny, 1995, and Kouider & Dupoux, 2009) is that there are two separate types of priming that contribute to repetition priming. Some of the priming effect comes from repeated access to the same lexical representation or access code, while some comes from repeated exposure to the same recent episode (i.e., the experience of hearing or seeing a particular physical stimulus). In other words, “repetition priming” is abstract lexical priming plus episodic priming. Morphological priming, on the other hand, involves facilitation from repeated activation of the same lexical representation or access code, but any facilitation deriving from repeated exposure to the same episode would be reduced or eliminated (since the actual experience of seeing, e.g., *assume* is different than the experience of seeing *assumption*).

If this explanation is correct, then the difference between repetition and morphological priming should disappear or at least be reduced when primes and targets differ in episodic details – e.g., when prime and target are written in different fonts or spoken in different voices. In that case, even when the prime and target are the same word, they aren’t the same physical experiences, and thus there will no longer be any reason for repetition priming to elicit a greater effect than morphological priming. Results on this question are mixed.

Consistent with the account presented above, Kouider and Dupoux (2009) found comparable long-lag repetition and morphological priming when the prime was spoken in a different voice than the target; this suggests that repetition priming is indeed the same thing as morphological priming, once episodic factors are ruled out (see, however, Wilder et al., 2019, for a slightly different explanation of this same set of results). This study is also one of the most similar to ours, since its targets were presented auditorily, as ours are, whereas most of the other studies reviewed here are visual.

Several other studies' results seem to conflict with Kouider and Dupoux's (2009) results, though. Bowers and Michita (1998) examined long-lag repetition and episodic priming in Japanese, which has multiple writing systems, by repeating the prime and target either in the same script (e.g., both in kanji, or both in hiragana) or in different scripts (kanji-hiragana or hiragana-kanji). While they do not report statistical comparisons between these conditions, their magnitudes of priming appear similar, contra Kouider and Dupoux (2009). Feldman and Moskovljević (1987) find a similar pattern with morphological and repetition priming in Serbo-Croatian, which can be written with two different alphabets. Brown and colleagues (1984) test a similar manipulation using Hindi and Urdu, but their results are ambiguous with respect to the issue at stake here: in their study, full repetition priming is numerically, but not significantly, larger than cross-script priming.

It may be the case that the extent to which episodic priming contributes to the difference between long-lag repetition and morphological priming is moderated by how idiosyncratic the differences between prime and target are. Differences between voices are relatively idiosyncratic, whereas differences between the same word written in, e.g., kanji vs. hiragana, or Latin vs. Cyrillic characters, may be more systematic. Furthermore, McLennan and Luce (2005; see also Luce et al., 2003) demonstrate that whether or not full repetition priming is larger than different-voice "repetition" priming is moderated by the speed, difficulty, and implicitness vs. explicitness of the experimental task; it is not obvious, though, that these factors can explain the differences between the sets of results described above.

### 1.3 The present study

The present study examines whether hearing one form of a Mandarin morpheme activates phonologically-derived variants of the same morpheme. As described above, a Mandarin syllable that has Low tone underlyingly is instead pronounced with something like a Rising tone, in certain phonological contexts. Thus we wanted to see whether hearing *shi<sup>L</sup>* (for example) activates *shi<sup>R</sup>*, and vice versa. Unlike all the previous studies on priming of phonologically-conditioned tonal variants, we use long-lag priming, which allows us to rule out potential contributions of meaning and form in order to identify priming that is purely due to repeated access of the same lexical representation. Note that, even though the change between Low and Rising tone only occurs in phonological contexts that license it, we tested single-syllable stimuli in isolation in order to avoid the complication of having potential priming from both syllables of compound words, not to mention the additional complication that it would not be possible to include Low tones in the same context as Rising tones (since Low tones would not survive in a context that triggers tone sandhi); the consequences of the use of single-syllable stimuli will be revisited in Section 5. Below we summarize the specifics of the three experiments in this study, and preview their findings.

In the first experiment, we compare the priming yielded by phonological variants of the same morpheme (e.g., *shi<sup>L</sup>* and *shi<sup>R</sup>*, which are both possible pronunciations of a morpheme whose underlying form is *shi<sup>L</sup>*) to priming yielded by form-related primes with no morphological relationship to the target. For example, *shi<sup>H</sup>* shares the same segments with *shi<sup>L</sup>* but has a different tone; thus, these words differ by one phoneme. Crucially, there is no systematic phonological relationship between Low and High tone in Mandarin (i.e., there is no phonological alternation that changes Low to High or vice versa), so *shi<sup>H</sup>* is not a possible pronunciation of *shi<sup>L</sup>*, nor vice versa. These two stimuli would thus be phonetically, but not morphologically, related. As discussed above, previous long-lag priming experiments in other languages suggest that there should not be priming for form-only overlap; however, we do not yet know if that is the case for tones, as all previous long-lag priming studies on form have used segmental or orthographic manipulations. Thus, in order to conclude that we are observing morphological priming, we should find facilitation of *shi<sup>L</sup>* by the prime *shi<sup>R</sup>*, or vice versa, but no facilitation by a morphologically unrelated prime like *shi<sup>H</sup>*.

This first experiment surprisingly failed to find priming in either condition. Thus, in order to make sure that long-lag priming works at all for Mandarin monosyllables, we follow it up with a second experiment including exact repetition priming (e.g., *shi<sup>L</sup>* primed by *shi<sup>L</sup>*), where facilitation should be expected under any account. In this experiment we successfully observe repetition priming – ruling out the possibility that this paradigm just doesn’t work with Mandarin, or that some problem with our experiment design prevented us from being able to detect priming effects – but still no morphological priming.

Finally, in a third experiment we test whether the repetition priming was due entirely to episodic priming, with no contribution from lexical access to the same lexical entry. We do this by testing “repetition” priming with the same word spoken in different voices, following Kouider and Dupoux (2009). Surprisingly, however, while repetition priming still holds (and with a similar magnitude as it did with identical voices in Experiment 2), morphological priming is still absent. Taken together, the results suggest that lexical repetition priming – beyond mere episodic priming – does occur in Mandarin monosyllables, but that there is no morphological priming of monosyllable allomorphs removed from any larger context. These results raise questions about what the underlying mechanisms for morphological priming are.

Materials (including recorded stimuli and experiment control scripts), data, and analysis scripts for all experiments are available at <https://osf.io/pt8qv>.

## 2. Experiment 1

Experiment 1 tested whether there would be greater long-lag priming when a word was primed by a neighbour with which it shared a morphological relationship via tone sandhi (e.g., *shi<sup>L</sup>* primed by *shi<sup>R</sup>*) than when it was primed by a neighbour with which it shared no morphological

relationship (e.g., *shi*<sup>L</sup> primed by *shi*<sup>H</sup>). Priming was measured by comparing each of these conditions' reaction times to the reaction time when the word was preceded by a completely unrelated prime (e.g., *shi*<sup>L</sup> preceded by *hua*<sup>F</sup>).

We predicted that the morphologically-related condition would elicit faster reaction times than the form-related condition and the unrelated condition because of long-lag morphological priming (e.g., Kouider & Dupoux, 2009). We did not expect the form-related condition to yield faster reaction times than the unrelated condition, since form relationship without morphological relationship typically does not engender long-lag priming. Nevertheless, we included this condition as a control, in order to rule out the possibility that any priming observed in the morphological condition is just based on the form relationship (while the previous long-lag priming literature already suggests that form relationships should not cause priming anyway, we could not be sure that that conclusion also applies to tone relationships since no study yet has tested those in long-lag priming).

If form priming does somehow occur in the present study, then a difference in reaction time between the morphologically-related and form-related conditions could be due to differences in phonetic similarity between the primes and targets. Specifically, primes and targets in the morphological condition consist of Low and Rising tones, which are the most phonetically similar pair of tones in standard Mandarin,<sup>4</sup> whereas primes and targets in the form condition consist of tones which are not phonetically similar. In other words, if there is more priming in the morphological condition than the form condition, this could be due not to the presence of a morphological relationship but just to the fact that the “morphologically-related” primes and targets also happen to be more phonetically similar (this, of course, requires assuming that long-lag form priming occurs in spoken Mandarin, unlike almost all other languages tested in previous studies). This confound is difficult to rule out in the present paradigm, given that comparing morphological vs. phonetic priming necessarily involves making comparisons across different pairs of tones.

As a more exploratory analysis, we also compare morphological priming in two directions: priming from Low-tone primes to Rising-tone targets, versus priming from Rising-tone primes to Low-tone targets. The Mandarin tone alternation described above is one-directional: underlyingly Low-tone morphemes are sometimes pronounced with something like a Rising tone, but underlyingly Rising-tone morphemes are never pronounced with Low tone. We did not have *a priori* hypotheses about how this difference would affect priming in a long-lag paradigm, but we report this comparison in case it helps generate future hypotheses.

---

<sup>4</sup> Low and Rising tones are easily confused with one another (Chuang & Hiki, 1972; Sharma et al., 2015; among others), the difference between them is less detectable in habituation paradigms (Chandrasekaran et al., 2007; among others), and they prime one another in implicit production priming (Chen & Chen, in press).



All methods for the experiment were pre-registered at <https://osf.io/d3xu5/>.

## 2.1 Methods

### 2.1.1 Participants

The final dataset includes data from one hundred fifty-two native speakers of Mandarin (average age 20.3, range 18–28, 63 men and 89 women; see <https://osf.io/6w8sy/> for detailed demographic information). An additional participant took part but did not press the appropriate keys during the experiment (on every trial this participant pressed an irrelevant key and waited until the trial timed out) and thus provided no meaningful data. One more participant took part but the data from this participant were excluded after the researchers found that the participant was under age 18 at the time of the experiment. Procedures for the experiment were approved by the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University. All participants provided written informed consent and were paid for their participation. While we had pre-registered a sample of 150 participants, the final sample slightly overshot this goal because participants attended the experiment in groups (see 2.1.3) and, in case of no-shows, we scheduled slightly more participants than needed. The decision to stop data collection was not influenced by the data themselves; rather, data collection stopped after the session in which the target sample size was reached.

### 2.1.2 Materials

The critical materials comprised 96 stimulus sets; one example stimulus set is shown in **Table 1**. Each set consisted of several existing Mandarin monosyllables. Two monosyllables served as possible target stimuli; these were segmentally identical and only differed in tone, with one carrying Low tone and one Rising tone. For the Low-tone target, a segmentally identical syllable with Rising tone served as the morphologically-related prime, and for the Rising-tone target a segmentally identical syllable with Low tone served as the morphologically-related prime. For each target, a segmentally identical syllable with either High or Falling tone served as the form-related (morphologically unrelated) prime, and a segmentally different syllable with either High or Falling tone served as the completely unrelated prime.

**Table 1.** Sample stimulus set.

	Low-tone target: <i>shǐ</i> <sup>L</sup>	Rising-tone target: <i>shǐ</i> <sup>R</sup>
Morphologically-related prime	<i>shǐ</i> <sup>R</sup>	<i>shǐ</i> <sup>L</sup>
Form-related prime	<i>shǐ</i> <sup>H</sup>	<i>shǐ</i> <sup>H</sup>
Unrelated prime	<i>hua</i> <sup>F</sup>	<i>hua</i> <sup>F</sup>

Forty-eight sets of existing Mandarin monosyllables were also used as fillers. With the set of critical stimuli described above, a Low or Rising tone in the primes block will always be followed by a segmentally matching syllable in a later target block, while the same is not true for High and Falling tones. To avoid having such a pattern over the course of the experiment, the filler sets had High- or Falling-tone targets, which could be preceded by a segmentally identical syllable with Falling or High tone (e.g., *jiao*<sup>H</sup> could be primed by *jiao*<sup>F</sup>, and *tang*<sup>F</sup> could be primed by *tang*<sup>H</sup>), or by a segmentally identical syllable with Low or Rising tone, or by a syllable with both different segments and different tones.

Finally, 288 monosyllables not attested in Mandarin were used as pseudoword foils for the lexical decision task. These syllables were all phonotactically legal. Some syllables were tonal gaps, i.e., syllables attested in other tones (e.g., *gao*<sup>R</sup>, which is not attested in standard Mandarin even though *gao*<sup>H</sup> is); others were accidental gaps, i.e., syllables not attested in any tone but also not violating any phonotactic constraint (e.g., neither *tei*<sup>H</sup>, *tei*<sup>R</sup>, *tei*<sup>L</sup>, nor *tei*<sup>F</sup> exist in standard Mandarin, even though there is no obvious general constraint barring the segmental sequence *tei* – e.g., *dei* is attested); see Gong & Zhang (2021) for further details on types of Mandarin pseudowords. These stimuli ensured a word-nonword ratio of 50% (96 critical targets, their 96 primes, plus 48 filler targets and their 48 primes, adds up to 288 words).

Several repetitions of each stimulus were produced in isolation, in a random order, by a male native speaker of Mandarin from Changchun. The first and second authors selected the clearest token of each stimulus and saved it into its own sound file.

### 2.1.3 Procedure

As the critical stimuli fell into six conditions, they were organized into six lists following a Latin Square design. (The 48 real-word fillers fell into three conditions, and were thus organized into three lists; thus, lists 1 and 4, for instance, had the same fillers.) The same pseudoword foils were used in all lists.

Experiment control and logging of responses and response times were handled by DMDX (Forster & Forster, 2003). To handle the long-lag priming, the stimuli were organized into blocks. For example, “block 1” consisted of six critical targets, three filler targets, the nine corresponding primes for these targets, and eighteen pseudo-syllables. Furthermore, these blocks were each divided into “prime” blocks and “target” blocks. For example, Prime Block 1 contained the primes for critical items 1–6, the primes for filler items 1–3, and nine pseudo-words; Target Block 1 contained the targets for critical items 1–6, the targets for filler items 1–3, and nine more pseudo-words. Two prime blocks were presented one after another, with the trials within a block fully randomized, and then two target blocks were presented one

after another, with the trials within a block fully randomized. So, for instance, the experiment began with the sequence Prime Block 1, Prime Block 2, Target Block 1, Target Block 2. Given that each of these “blocks” included 18 trials, this means that the minimum number of trials intervening between a target and its prime was 18 (if the trials were randomized such that, for example, item 5’s prime occurred at the very end of Prime Block 1 and its target at the very beginning of Target Block 1, with the 18 trials of Prime Block 2 intervening) and the maximum number of trials intervening between a target and its prime was 52 (if, for example, item 5’s prime occurred at the very beginning of Prime Block 1 and its target at the very end of Target Block 1, such that in between there were the 17 remaining trials of Prime Block 1, all 18 trials of Prime Block 2, and the 17 remaining trials of Target Block 2). Thus, lag between prime and target followed a roughly normal distribution constrained between 18 and 52 trials. After each sequence of two prime blocks and two target blocks, the participant was given an opportunity to rest; thus, the experiment was overall divided into eight sections, with seven rests in between. The “blocks” described above were a hidden part of the experiment control design and were not known to participants; from the participant’s point of view, they simply experienced 72 trials continuously, followed by an opportunity for a break, and then repeated that process seven more times.

On each trial, a “###” fixation mark was shown at the center of the screen for approximately 600 milliseconds (36 frames, on a monitor with a 60 Hz refresh rate), and then the auditory stimulus was presented over headphones. Participants made a lexical decision to every trial, including primes. They were instructed to press the right shift button if the sound they heard was an existing syllable of Mandarin, or the left shift button otherwise. They were instructed to respond as quickly and accurately as possible. Reaction times were recorded relative to stimulus onset.

Twenty practice trials preceded the experiment proper. Participants were tested in a quiet computer lab at the Hong Kong Polytechnic University in groups of three to eighteen. All communication with the participants was carried out in Mandarin.

#### **2.1.4 Analysis**

Based on our *a priori* pre-registration, responses that were incorrect, faster than 200 ms, or more than 1.5\*IQR above or below the corresponding participant median or item median were removed. The visualizations and statistical models were carried out only on the critical Low- and Rising-tone targets remaining after these exclusions.

We visualized the data using informative plots motivated by the experimental design (Politzer-Ahles & Piccinini, 2018) and supplemented these visualizations with mixed-effect models (Baayen et al., 2008). Response time was regressed on critical predictors and nuisance

covariates using the following model, implemented in the `{lme4}` package (Bates et al., 2015) of the R statistical computing environment (R Core Team, 2016):

```
RT ~ (form+control)*TargetTone + zLag + zStimDuration + zTrialNumber
+ (1|Participant) + (form+control||Item)
```

The effect of prime type was modeled as two dummy variables, “form” and “control”, each representing the difference between the morphologically-related condition and the form-related or unrelated (control) condition. This is equivalent to dummy-coding (treatment-coding) one “PrimeType” variable with the morphological condition as the reference level. This allows us to directly compare the morphologically-related condition to each other condition. The use of two manually-coded dummy variables (rather than one “PrimeType” factor) is an R hack that, along with the double-pipe syntax in `lme4` (“||”), allows us to suppress correlations between random effects.

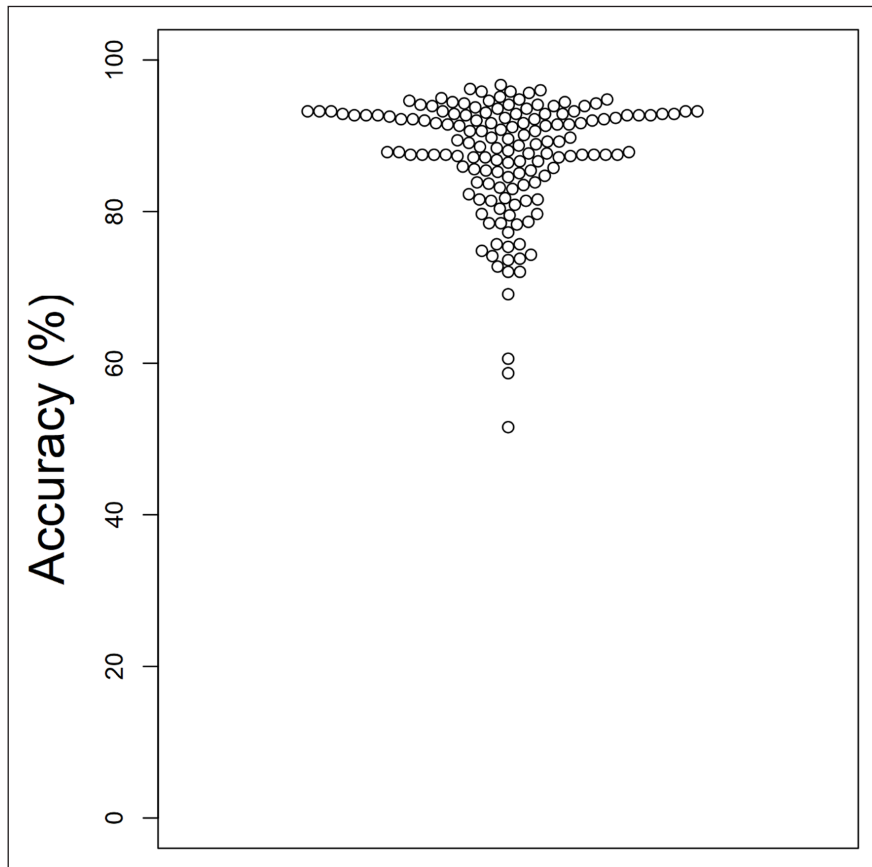
TargetTone refers to whether the target was a syllable with Low or Rising tone. This variable was deviation-coded with Low tone as the baseline level (i.e., Low tone was coded at -0.5 and Rising tone as 0.5) so that coefficients for PrimeType would represent main effects across levels of TargetTone.

The remaining fixed effects are nuisance covariates included to reduce unexplained variance in the model. Lag refers to the number of trials intervening between the target and its corresponding prime; StimDuration refers to the duration of the auditory stimulus; and TrialNumber refers to the ordinal number of the trial in the course of the experiment. Each of these factors was z-scored before analysis to make the coefficients of theoretical importance easier to interpret and to help model convergence. Finally, random effects of the fixed factors of interest were included for participants and items. Since the model was very complex, random effects for nuisance covariates were not included (see, e.g., Barr et al., 2013), nor were correlations between random slopes and random intercepts. As the model with this large number of parameters did not converge within the default iteration limit, the model was re-run with the limit set to 50,000 iterations and with the BOBYQA optimizer. Attempting to fit the full random effects structure that we pre-registered resulted in singular fit, so we simplified it to the random effects structure described above by removing random effects with zero or near-zero variance.

## 2.2 Results

As shown in **Figure 1**, accuracy among most participants was high, although three participants had very low (near chance) accuracy. As we did not pre-register any exclusions based on

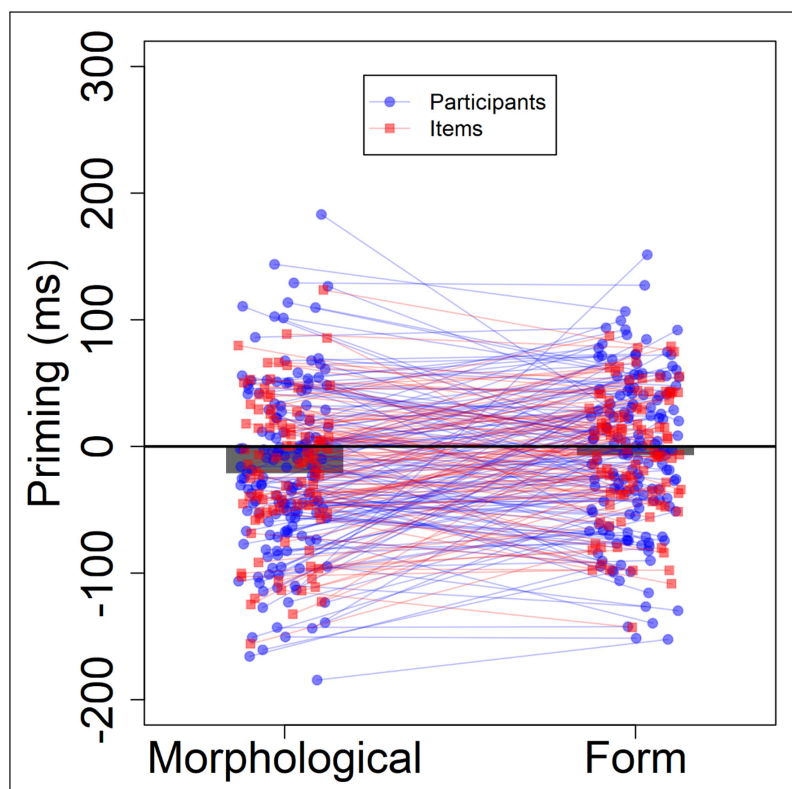
participant-wise accuracy, and the sample is large enough that the results are unlikely to be skewed by a few participants' performance, we did not exclude these participants.



**Figure 1:** Beeswarm plot (i.e., a univariate scatterplot with the dots shifted horizontally to avoid overlapping, with the added effect that the plot roughly simulates a violin plot) showing the distribution of accuracy scores across participants. Each dot represents the accuracy (across all trials, including primes, fillers and pseudowords) for one participant. The vertical axis represents participant accuracy, ranging from 0% to 100%. The majority of dots are clustered high on the graph, between 80% and 100%, with only a few participants below this range (particularly, three participants with accuracies near or below 60% stand out from the rest of the distribution).

**Figure 2** shows the priming effects to the critical targets. The left side shows the priming effects for the morphologically-related condition, and the right side the form-related condition. Priming effect size (unrelated minus morphologically-related, or unrelated minus form-related) is represented on the vertical axis, such that facilitative priming effects should be above zero. Each dot represents one participant's or item's priming effect in the given condition, with a line connecting a given participant's/item's dot in the morphological condition to the corresponding

participant's/item's dot in the form condition. These dots show the general distribution of the priming effects in each condition; in fact, for each condition, the dots are roughly evenly scattered above and below zero (rather than mostly above or mostly below zero), suggesting that there is not a robust priming effect for either condition. The lines show the general distribution of the relationship between the conditions; since there is a roughly even mixture of upward- and downward-sloping lines (rather than most of the lines sloping upward or most of the lines sloping downward), this suggests that there is not a robust difference between the morphological and form priming effects. For ease of interpretation, gray bars are also shown underneath the clouds of dots; the heights of the bars indicate the mean priming effect for each condition. Each bar is below zero; that is to say, both morphologically-related and form-related targets elicited *slower* response times than totally unrelated targets. This suggests that, contrary to our hypothesis, morphological relatedness did not elicit a facilitative priming effect. Since neither condition elicited priming (and if anything, the morphological primes facilitated responses *less* than the form primes did), the confound with phonetic similarity (i.e., the fact that the morphologically-related primes are more phonetically similar to their targets than the form-related primes are) is moot.



**Figure 2:** Priming effects in Experiment 1. See main text for details.

While the visualization in **Figure 2** strongly suggests that the results do not support our hypothesis, we also ran a mixed-effects model on the reaction time data to account for the influence of nuisance covariates that may not be reflected in this visualization. The results of the model (see 2.1.4 for details of the model specification) were consistent with the description above, and are summarized below.

The nuisance covariates yielded reasonable coefficients in the model, which can also function as a quality check to ensure the data are behaving as auditory lexical decision data ought to (note that the parameter estimates below have been transformed from z-scored units back into original units): the lag between prime and target was not significantly associated with reaction time ( $b = -0.41, t = -0.21$ ), but longer stimuli were associated with substantially longer reaction times ( $b = 56.60, t = 13.82$ ), and later trials in the experiment were responded to more quickly ( $b = -15.09, t = -1.87$ ).

Most importantly, contrary to our hypothesis, the form-related condition was responded to 13ms more quickly than the morphologically-related condition ( $b = -13.35, t = -2.75$ ), and the unrelated condition 21ms more quickly than the morphologically-related condition ( $b = -20.74, t = -4.22$ ). As we only had predictions about facilitative priming for the morphologically-related condition, and the effect observed here is in the opposite direction, we do not make any claims about statistical significance of this unexpected inhibition pattern; rather, we test it in a confirmatory replication in Experiment 2.

The other effect of interest is the interaction between prime type and target tone. The interaction was barely significant in an omnibus test using model comparison ( $\chi^2(2) = 6.40, p = .041$ ). While the effect of prime type, as discussed above, was opposite the expected direction, this pattern was stronger in Rising-tone targets than in Low-tone targets. Specifically, for Rising targets, the form-related condition was 21ms faster than the morphological condition ( $b = -20.51, t = -2.96$ ) and the unrelated condition 33 ms faster ( $b = -32.82, t = -4.74$ ); for Low targets, though, the form-related condition was only 6ms faster than morphological ( $b = -6.19, t = -0.91$ ) and the unrelated condition only 9ms faster ( $b = -8.66, t = -1.27$ ).

### 2.3 Discussion

Contrary to our expectation, morphologically-related prime-target monosyllable pairs in Mandarin did not trigger facilitative priming; in fact, morphologically-related targets were responded to more slowly than unrelated targets. This is different than what has been observed for long-lag morphological priming in other experiments.

It would be premature, however, to conclude from this experiment that morphological relationships based on Mandarin tone work differently than morphological differences based on segmental alternations in other languages. Instead, it is possible that the long-lag priming



paradigm simply does not work with Mandarin monosyllables, for whatever reason – for instance, maybe the experiment only works with multisyllabic stimuli like those used by Kouider and Dupoux (2009), or maybe there are some other characteristics of Mandarin that we had not considered but which cause this paradigm to not work. Therefore, to test whether the paradigm works at all, it was necessary to run the experiment with an identity priming condition as a manipulation check (e.g., *shi*<sup>L</sup> primed by *shi*<sup>L</sup>, the exact same stimulus). If the long-lag priming paradigm works with Mandarin monosyllables, this condition should elicit faster reaction times than the unrelated condition, regardless of the psychological status of Mandarin tone alternation. On the other hand, if our failure to observe morphological priming in the above experiment was because long-lag priming just doesn't work in Mandarin, then this identity priming condition should also not elicit faster reaction times than the unrelated condition.

The second reason for the new experiment was to test the unexpected inhibition effect. As we had predicted a facilitative effect, we are not in a position to make conclusions about the reality of the inhibition for morphologically-related pairs observed in Experiment 1; that is to say, it is possible that this pattern was just noise and that claiming there is inhibition would be a Type I error. We can only say that we did *not* find evidence for facilitative priming, not that we did find strong evidence for inhibition. Therefore, in Experiment 2 we again included the comparison between morphologically-related and completely unrelated, in order to test this pattern in a confirmatory way now that it could be explicitly predicted.

### 3. Experiment 2

As described above, the purpose of Experiment 2 was to test whether Mandarin monosyllables trigger identity priming, and to test whether the inhibition effect for morphologically-related pairs compared to unrelated pairs would be replicated. Therefore, this experiment included three priming conditions: identical, morphologically-related, and unrelated. We discarded the form-related condition (although this manipulation was still present in the fillers); since the previous experiment did not find morphological priming at all, it was no longer necessary to try to disentangle morphological priming from form priming. Furthermore, for this experiment we were no longer interested in the comparison between Low-tone and Rising-tone targets; nevertheless, we still report these comparisons, in the interest of testing whether the unexpected asymmetry in Experiment 1 is replicable.

#### 3.1 Methods

##### 3.1.1 Participants

One hundred fifty-three native speakers of Mandarin (average age 23.9, range 19–35, 32 men and 121 women; see <https://osf.io/f8tb9/> for detailed demographic information). This sample

size was chosen to match the sample size of Experiment 1 (which had 153 participants before one was removed; see 2.1.1). Procedures for the experiment were approved by the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University, and all participants provided written informed consent and were paid for their participation.

### 3.1.2 Materials

The critical materials comprised 72 stimulus sets. Each item only occurred with one target tone, Low or Rising. Each of the 72 targets could appear with either an identical prime (e.g., *shi*<sup>L</sup> preceded by *shi*<sup>L</sup>), a morphologically-related prime (e.g., *shi*<sup>L</sup> preceded by *shi*<sup>R</sup>), or an unrelated prime (e.g., *shi*<sup>L</sup> preceded by *hua*<sup>F</sup>).

There were several types of fillers. 24 fillers consisted of Low- or Rising-tone targets preceded by a prime with the same segments but High or Falling tone (analogous to the form-related condition in Experiment 1). An additional 48 fillers consisted of targets with High or Falling tone, and priming conditions analogous to those in the rest of the experiment (12 of these targets were preceded by identical primes, 12 by unrelated primes, 12 by phonetically-related primes with the same segments but either Low or Rising tone, and 12 by phonetically-related primes with the same segments but either High or Falling tone – whichever mismatched the target). Finally, the same 288 monosyllables from Experiment 1 were used as pseudoword foils.

Forty-six of the syllables used in Experiment 2 were new, and were recorded by the same speaker as the stimuli in Experiment 1. The remaining syllables were the same as in Experiment 1, and the same recordings were re-used.

### 3.1.3 Procedure

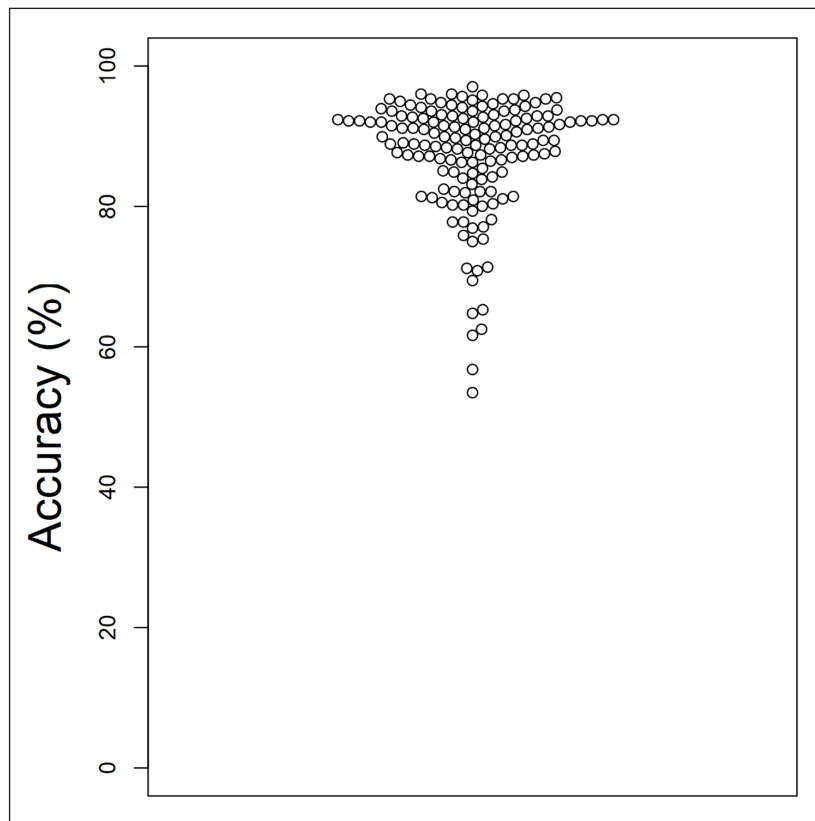
As the experiment had three conditions rather than six, the critical stimuli were organized into three Latin Square lists. Fillers and pseudowords were the same across all lists. The procedure was otherwise identical to that of Experiment 1.

### 3.1.4 Analysis

Data preprocessing was carried out as in Experiment 1. The modeling strategy was also the same, except that the unrelated condition was coded as the baseline so that both the identical and morphologically-related conditions could be directly compared to this one; thus, the effect of prime type was implemented as two dummy variables, “ident” and “morph”, each comparing the corresponding condition to the unrelated condition. The statistical model was thus  $RT \sim (\text{ident} + \text{morph}) * \text{TargetTone} + z\text{Lag} + z\text{StimDuration} + z\text{TrialNumber} + (1 | \text{Participant}) + (\text{ident} + \text{morph} || \text{Item})$ .

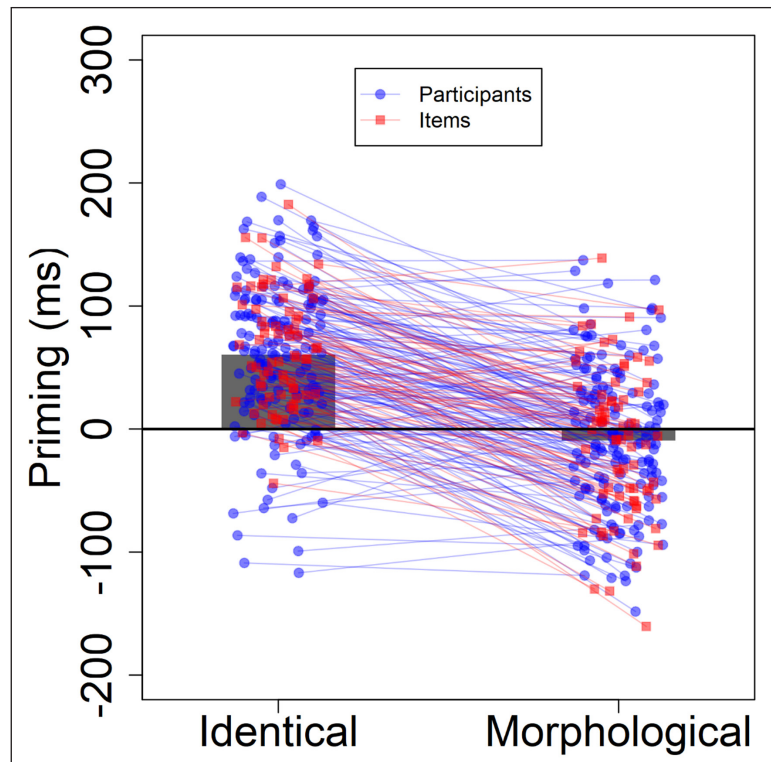
### 3.2 Results

As in Experiment 1, accuracy was fairly high for most participants, as shown in **Figure 3**.



**Figure 3:** Beeswarm plot showing the distribution of accuracy scores across participants. As in Experiment 1, the majority of participants are between 80% and 100%, with only a few participants below this range.

The critical priming effects are shown in **Figure 4**, using the same type of graph used for the previous experiment. In this case the left side shows the priming effects for the identity condition, and the right side the morphologically-related condition. In this case we can see that for the identity condition, most of the dots are above zero, suggesting that there is a fairly reliable priming effect (i.e., that unrelated is slower than identity for most participants and items); on the other hand, for the morphological condition, the dots are roughly evenly distributed above and below zero, as in Experiment 1, suggesting that there is little priming effect. Furthermore, the majority of the lines appear to be sloping downward, suggesting that the identity priming effect is bigger than the morphological priming effect for most participants and items. Finally, the bars show that the identity priming effect is fairly large (about 60 ms), whereas the morphological effect is a small inhibition effect.



**Figure 4:** Priming effects in Experiment 2. See main text for details.

The statistical model provided further support for this pattern. Targets preceded by identical primes were responded to 64ms faster than targets preceded by unrelated primes ( $b = -63.90$ ,  $t = -13.51$ ), whereas targets preceded by morphological primes were responded to 10 ms more slowly than targets preceded by unrelated primes ( $b = 10.32$ ,  $t = 1.88$ ); the latter effect is significant in a one-tailed test, based on our *a priori* expectation that there would be inhibition like there was in Experiment 1.

On the other hand, it is worth mentioning that the nuisance covariates did not show the same effects as observed in Experiment 1. Longer stimulus duration was associated with a small and non-significant increase in reaction time ( $b = 9.08$ ,  $t = 0.63$ ), while later trials were associated with a non-significant decrease in reaction time ( $b = -13.05$ ,  $t = -0.92$ ). Finally, targets with more lag between themselves and their primes were associated with marginally slower reaction times ( $b = 3.37$ ,  $t = 1.78$ ). These results are roughly opposite those of Experiment 1, where lag was not a significant predictor of reaction time but stimulus duration and trial number were.

The interaction between priming condition and target tone was again significant ( $\chi^2(2) = 7.40$ ,  $p = .025$ ). Specifically, while the size of the identity priming effect did not significantly differ between Low- and Rising-tone targets ( $b = -6.42$ ,  $t = -0.67$ ), the size of the morphological priming effect was 30 ms smaller with Low targets than it was with Rising targets ( $b = -29.90$ ,  $t = -2.72$ ). While Rising-tone targets were significantly inhibited by Low-tone morphological

primes (25ms slower than unrelated:  $b = 25.27$ ,  $t = 3.22$ ), Low-tone targets were not significantly inhibited by Rising-tone morphological primes, and in fact were numerically, but not significantly, faster than Low-tone targets with unrelated primes ( $b = -4.63$ ,  $t = -0.60$ ). This mirrors the pattern observed in Experiment 1, in which Rising targets were also inhibited by morphological primes but Low targets were not.

### 3.3 Discussion

This experiment demonstrated that the long-lag priming paradigm does indeed work with Mandarin monosyllables: identity priming, the most basic manipulation check, does occur (targets preceded by identical primes are responded to faster than targets preceded by unrelated primes). Thus, the failure to observe morphological priming in Experiment 1 was not because the paradigm overall does not work; it must be due to other reasons.

This experiment also confirmed the unexpected morphological inhibition pattern found in Experiment 1; while this pattern was an unpredicted surprise in that experiment, here we observed it when it had been explicitly predicted. Together, these experiments suggest that this particular morphological relationship causes inhibition, rather than facilitative priming, in Mandarin long-lag priming with auditory lexical decision.

This pattern of results leaves us with a conundrum. In some models of morphological representation, a morphological relationship should be the same as an identity relationship, as both involve accessing the same lexical representation twice (Stockall & Marantz, 2006; Kouider and Dupoux, 2009). Why, then, do we not observe morphological priming when identity priming is possible? One potential explanation, of course, is that this assumption about morphological representation is wrong (i.e., morphological priming is not the same thing as identity priming). Another, however, is that the priming observed here was not linguistic priming at all (i.e., not based on activation of a lexical or phonological representation), either identity or morphological priming, but rather was *episodic priming*: recognition of the same physical event that happened recently. This recognition could occur without accessing linguistic representations at all. Kouider and Dupoux (2009) demonstrated such a pattern in French. With relatively short lags between prime and target, and when identical primes and targets were the same physical stimuli, identity priming was larger than morphological priming; on the other hand, with a longer lag between prime and target, and when the “identical” prime and target were the same word spoken by different speakers (i.e., different physical stimuli), then identity priming was reduced to be the same size as morphological priming. Kouider and Dupoux argue that this occurred because the priming effect observed in the former situation (short lag, physically identical stimuli) was a sum of two distinct priming effects: repeated activation of a lexical entry or phonological representation (linguistic priming), and recognition of a familiar event (episodic priming). In the latter condition (long lag, physically different stimuli), the contribution of episodic priming was removed, leaving only the linguistic/morphological contribution, which was of the same

magnitude as it was in the identity priming condition. Under this understanding of contributions to long-lag priming, it is possible that our Experiment 2 still did not demonstrate that long-lag linguistic priming works in Mandarin monosyllables. Rather, it is possible that the identity priming effect we observed was wholly episodic in nature, rather than linguistic. In short, maybe there really is no linguistic long-lag priming effect for Mandarin monosyllables.

To test this, we performed a third experiment attempting to remove the contribution of episodic memory.

## **4. Experiment 3**

Following Kouider and Dupoux (2009), we used prime and target stimuli in different voices, spoken by different speakers, in order to test whether the identity priming effect observed in Experiment 2 was in fact an episodic priming effect. In all other respects the conditions were the same as in Experiment 2. If the long-lag priming paradigm with auditorily presented Mandarin monosyllables does not yield linguistic priming, we expect that neither the identical nor the morphologically-related conditions will elicit faster reaction times than the unrelated condition. On the other hand, if Mandarin monosyllables really trigger long-lag identity priming that is linguistic in nature, then the pattern observed in Experiment 2 should be replicated. In addition, Experiment 3 also served as an additional confirmatory test of whether the morphological condition would elicit inhibition relative to the unrelated condition.

### **4.1 Methods**

#### **4.1.1 Participants**

One hundred fifty-three native speakers of Mandarin (average age 23.5, range 19–36, 20 men, 132 women and one who preferred not to indicate the gender; see <https://osf.io/mnb8f/> for detailed demographic information). This sample size was chosen to match the sample size of Experiment 1 (which had 153 participants before one was removed; see 2.1.1). Procedures for the experiment were approved by the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University, and all participants provided written informed consent and were paid for their participation.

#### **4.1.2 Materials**

The stimulus design was the same as in Experiment 2. In addition, the same materials were re-recorded by a female Mandarin native speaker from Harbin, yielding two recordings of each stimulus. In all other respects the materials were the same as in Experiment 2.

#### **4.1.3 Procedure**

While this experiment had the same three conditions as Experiment 2, each condition could be realized with the male voice as the prime and female voice as the target, or vice versa. Thus, we

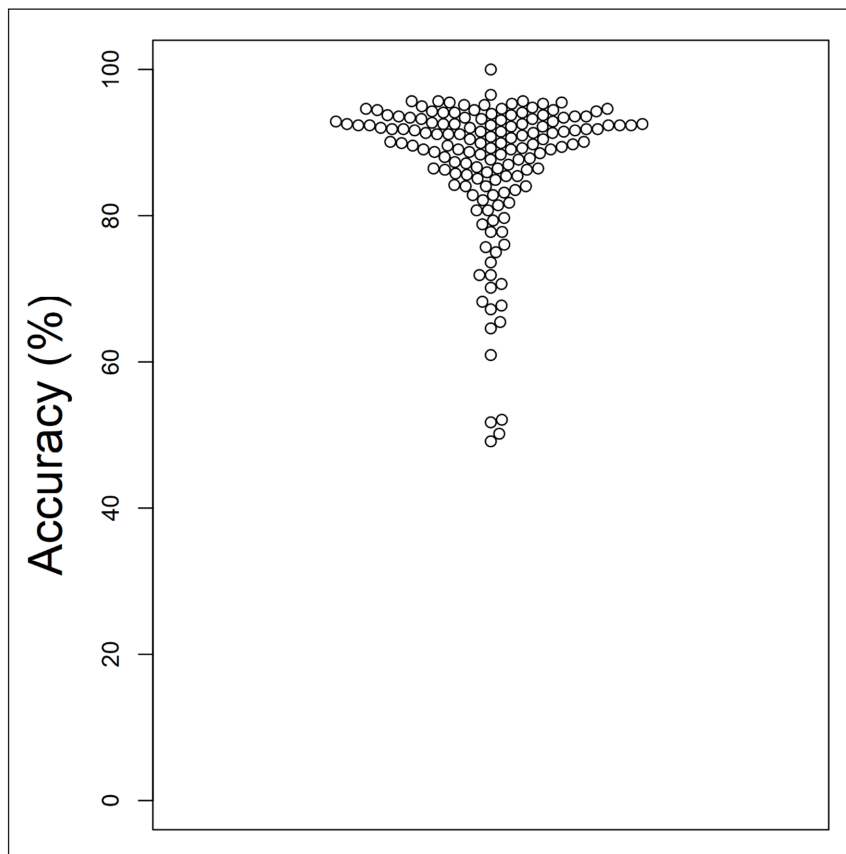
divided the critical stimuli into six Latin Square lists, treating the target's voice as a between-participants factor. In all other respects the procedure was the same as in Experiment 2.

#### 4.1.4 Analysis

The analysis was the same as in Experiment 2.

#### 4.2 Results

Accuracy was fairly high again, as shown in **Figure 5**. Two participants had accuracy near zero, which is likely due to these participants mixing up which key to press for words and which to press for pseudowords; therefore, we switched the correct/incorrect labels in these participants' data for the analysis (recall that trials with incorrect responses are removed in our analysis).

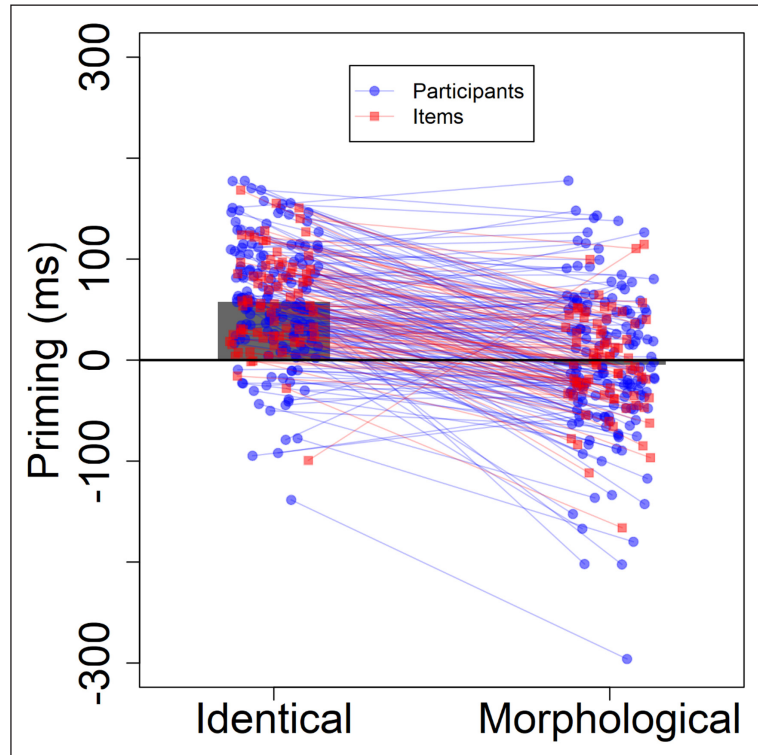


**Figure 5:** Beeswarm plot showing the distribution of accuracy scores across participants.

The critical priming effects are shown in **Figure 6**, using the same type of graph used for the previous experiments. The results from this experiment look almost identical to those of Experiment 2: there is large priming for the identity condition and small inhibition for the morphologically-related condition, and the difference between these patterns is fairly robust



across participants and items. In other words, using different voices for the primes and targets had no appreciable impact on the results; the present experiment using different voices showed the same pattern as Experiment 2 using identical voices across primes and targets.



**Figure 6:** Priming effects in Experiment 3. See main text for details.

The statistical model confirmed what can be seen in the figure. There was a significant 59ms identity priming effect ( $b = -59.08$ ,  $t = -13.15$ ). Morphological inhibition, while not significant this time, was in the same direction as the previous experiments, with targets preceded by morphological primes responded to 7ms more slowly than targets preceded by unrelated primes ( $b = 7.26$ ,  $t = 1.40$ ).

The nuisance covariates followed the pattern of Experiment 1 and not that of Experiment 2: later trials were significantly faster ( $b = -19.01$ ,  $t = -2.09$ ), longer stimuli elicited longer reaction times ( $b = 35.32$ ,  $t = 7.71$ ), and prime-target lag did not substantially impact reaction times ( $b = -0.87$ ,  $t = -0.47$ ).

Unlike in Experiments 1 and 2, the interaction between priming condition and target tone was not significant this time ( $\chi^2(2) = 2.96$ ,  $p = .227$ ). The size of the identity priming effect did not significantly differ between Low and Rising targets ( $b = -7.11$ ,  $t = -0.79$ ). The size of the morphological priming effect was 17ms smaller with Low targets than it was with Rising

targets ( $b = -17.73$ ,  $t = -1.72$ ), but this is less of a difference than what was observed in the previous experiments. Specifically, Rising-tone targets were significantly inhibited by Low-tone morphological primes (16ms slower than unrelated:  $b = 16.12$ ,  $t = 2.18$ ), and Low-tone targets were not significantly inhibited by Rising-tone morphological primes ( $b = -1.61$ ,  $t = -0.22$ ). While this pattern did not yield a significant interaction, it is in the same direction as the pattern observed in Experiments 1 and 2.

### 4.3 Discussion

We again failed to find morphological priming, replicating the findings of Experiments 1 and 2. On the other hand, identity priming persisted, even when the primes and targets were presented in different voices. This suggests that the identity priming effect was not merely an episodic memory effect, but was a genuine lexical effect based on repeated access of the same lexical or phonological representation. The results, therefore, suggest that participants hearing, e.g., *shi<sup>L</sup>* activate its lexical entry (as evidenced by the identity priming for physically different tokens of *shi<sup>L</sup>-shi<sup>L</sup>*), but do not activate a lexical entry associated with its phonological variant *shi<sup>R</sup>* (as evidenced by the lack of morphological priming for *shi<sup>L</sup>-shi<sup>R</sup>*).

An alternative explanation could be that our identity priming effect *was* indeed episodic and the lag between prime and target was simply not enough to wipe out episodic priming. 18 to 52 trials intervened between prime and target in this experiment; by comparison, in Kouider and Dupoux's (2009) experiment which successfully eliminated episodic priming, there were 96–192 intervening trials. We are not aware of any study testing different-voice identity and morphological priming with a lag of about 18–52 trials, so we cannot be certain that this lag should be enough to eliminate episodic priming. Importantly, however, Kouider and Dupoux also found that same-voice identity priming diminished with increasing lags (82 ms at short lags, 62 ms at medium lags, and 47 ms at long lags with different voices), whereas morphological priming remained constant across lags and voices. In the present experiment, different-voice identity priming (60 ms) was similar in magnitude to same-voice identity priming in the previous experiment (64 ms). Our lag of 18–52 is also fairly typical, if not longer than average (albeit admittedly shorter than that of Kouider and Dupoux's different-voice experiment), as long-lag priming experiments go; for example, in a study by Wilder and colleagues (2019), a lag of just five intervening trials was enough to eliminate most of the difference between identity and morphological priming. Finally, given that the prime and target are physically different stimuli, it is unclear how the priming effect observed here *could* be episodic without some sort of abstraction – indeed, episodic priming should be ruled out of the present experiment by design. For these reasons, we find it more likely that the identity priming in the present experiment reflects genuine lexical access, rather than episodic memory.

## 5. General discussion

Across three experiments, we failed to find long-lag priming between different phonological variants of the same morpheme in Mandarin. This lack of priming cannot be explained by assuming that long-lag priming simply doesn't work in Mandarin, because it does – our experiments confirmed that long-lag identity priming occurs in spoken Mandarin stimuli and cannot be attributed to mere episodic effects. We are left, then, with the counterintuitive finding that Mandarin morphemes can prime themselves – even when pronounced in a different voice, so that they are not the same physical stimulus – but that they don't prime systematic phonological variants of themselves.

### 5.1 Comparison to previous studies of priming across allophonic variants

While we have described the present study's failure to observe morphological priming as surprising, it is arguably not unprecedented. In particular, the results of the present study are somewhat similar those of at least two previous studies.

The first is that of McLennan and colleagues (2003), who examined priming by allophonic variants based on American English flapping. In this variety of English, intervocalic alveolar stops preceding unstressed vowels, such as the /d/ in *raider* or the /t/ in *rater*, are often pronounced as alveolar flaps [ɾ]. McLennan and colleagues found that, under limited circumstances, tokens pronounced with the flap (i.e. [ɹeɪɾə]) fail to prime hyperarticulated tokens pronounced with full stop consonants [ɹeɪdə] or [ɹeɪtə], and vice versa. In both this study and our study, we see that one form of a word does not prime another form of the same word. Beyond this similarity, though, there are several important differences between these studies.

Most importantly, the conditions in which we did not find priming across different forms were precisely the conditions in which McLennan and colleagues predict it would be found. Specifically, they did not find across-form priming when participants' task in response to primes and targets was "easy lexical decision" (lexical decision in which the nonwords were not very word-like), whereas they did find across-form priming when participants performed shadowing or "hard lexical decision" (lexical decision in which the nonwords were more word-like) for the primes and/or the targets. They explain this pattern by proposing that abstract effects (i.e., priming across different forms of the same word) take time to emerge. Therefore, faster and easier lexical decision tasks tap into form-specific encoding that occurs before abstract underlying forms have been accessed (or, more specifically, before those underlying forms have established resonances with other possible surface forms of the same word), whereas more abstract encoding is tapped into by slower and harder lexical decision tasks, and by speech shadowing, which they argue requires deeper phonological processing than easy lexical decision does. If our study's lexical decision task was easy enough to allow fast processing that taps into more specific surface forms as opposed to abstract underlying forms, then we should have also found little to no

priming when the same word was presented in different voices; to the contrary, though, we found substantial (and not substantially reduced) priming in that situation (see also Luce & Lyons, 1998, who also find this pattern). In fact, as our nonword stimuli included a mix of more and less word-like stimuli (but even the least word-like stimuli were still phonological neighbours of real words), our task was arguably a “harder” lexical decision task and thus would be expected to have yielded abstract priming across different allophonic forms under their account.

The other situation in which McLennan and colleagues (2003) failed to find priming across different forms was when the same word was pronounced in different ways that do not trigger allophonic variation or ambiguity. Pronouncing *raider* casually as [ɹeɪrə] leads to ambiguity (the [r] could correspond to either a /d/ or a /t/), but pronouncing, e.g., *bacon* casually does not lead to any ambiguity (a fast, casual pronunciation of the /k/ in *bacon* still does not correspond to any phoneme other than /k/). While McLennan and colleagues (2003) found that casual, flapped tokens of *raider* prime hyperarticulated tokens of *raider* and vice versa, they also found that casual tokens of *bacon* did not prime carefully produced tokens of *bacon* or vice versa (see also McLennan et al., 2005). They explained this as resulting from ambiguity: specifically, they argue that hearing an ambiguous token causes listeners to activate multiple possible underlying forms and to subsequently “restore” their possible surface forms, whereas hearing an unambiguous form just activates the corresponding underlying form and then resonance between that underlying form and the initially perceived surface form reinforces activation of that same surface form (along with all its specific acoustic characteristics), without any need to activate other forms. For that reason, ambiguous forms trigger abstract priming whereas unambiguous forms trigger form-specific priming. This cannot explain the present study’s pattern of results, however, because our primes and targets were highly ambiguous (see further discussion of this point below) and the surface tone itself is also ambiguous in half of them (as a surface Tone 2 can correspond to underlying Tone 2 or Tone 3) and yet did not trigger abstract priming across different allophonic forms of the same morpheme, or across different allotones of the same underlying Tone 3. The main difference between the types of ambiguity in these two studies is that in our study the ambiguous tones correspond to the canonical form of one representation (Tone 2 is the canonical form of underlying Tone 2 and is the sandhi form of underlying Tone 3), whereas in McLennan and colleagues’ (2003, 2005) studies the flap is arguably not the canonical form of either /d/ or /t/, under traditional phonological theory. Perhaps, then, these studies’ differing results could be reconciled by positing that the speech comprehension algorithm first compares input to its canonical underlying form, and successful matching with a canonical underlying form precludes any attempt to match with other underlying forms (see also Sumner & Samuel, 2005, who argue that the lexical decision task “bolsters a preference for the canonical form” [p. 333]) – if so, then between-form priming might occur for other forms of tone alternation that yield a surface form which does not closely match any underlying form. This possibility remains to be tested.

All in all, the fact that our priming effect was not specific to individual speakers (i.e., priming persisted across indexical variability) but was specific to tone categories (i.e., priming did not persist across allophonic variants of the same morpheme) is difficult to reconcile with McLennan and colleagues' (2003) account; furthermore, the task which failed to elicit abstract priming in our experiments is arguably most similar to one of the tasks that did elicit abstract priming in theirs, and the sort of variation tested in our experiment is arguably more “abstract” and less “indexical” than the sort tested in theirs. Therefore, we cannot conclude that the lack of morphological priming observed in our study is demonstrating the same phenomenon, or has the same locus, as the lack of abstract priming they describe. For the same reason, we cannot assume that our lack of priming is predicted by the lack of priming in their study. Instead, we consider the lack of priming in our study to be unexpected based on existing theories of priming. The question that remains is why we failed to find priming under these circumstances.

Another study reporting results that appear similar to ours is one by Sumner and Samuel (2005), who found that word-final variants of /t/ in American English all appear to contact the same underlying representation in immediate priming (e.g., when there is no lag, *flute* primes *music* regardless of which allophone of /t/ is produced in *flute*) but not in long-lag priming. In long-lag priming, they argue that the carefully articulated canonical form of *flute* primes itself, but other forms do not prime themselves (e.g., *flute* with an unreleased [t̚] does not prime itself) and under no circumstances does one form prime another form. While our results are somewhat different, in that both variants prime themselves in our study (Rising-Rising identity priming was 59 ms and 53 ms faster than control in Experiments 2 and 3, respectively, and Low-Low identity priming was 60 and 62 ms faster than control), our results are similar to theirs in that we also failed to find priming across different variants at long lags. Sumner and Samuel's (2005) long-lag paradigm, however, is not comparable to ours, because they did not include unrelated controls for their critical targets; while they showed that different variants prime each other less than identity priming (e.g., the reaction times for a canonical [t] word primed by an unreleased [t̚] word are slower than those for a canonical [t] word primed by another canonical [t] word), they did not show that these variants don't prime each other at all (i.e., they did not compare [t̚]-[t] priming to unrelated priming within the same items; they did include a separate set of items with unrelated control primes, but this makes the crucial comparison between related and unrelated primes a between-items comparison). Therefore, it is difficult to be sure that this study did not observe across-variant priming whatsoever; if it did, however, that would be consistent with our own finding.

On the other hand, in a more recent study, Sumner (2013) found evidence that different variants *are* activated in long-lag priming. Specifically, careful and casual articulations of *center*, i.e. with or without a pronounced [t], both make it hard for participants to later reject a pseudoword *senner*. If the reason this occurs is because both forms activate *center*, then these results would be evidence for priming across different variants, contra the present study and the other studies discussed above that failed to find such priming.

## 5.2 Possible explanations for the lack of morphological priming

### 5.2.1 Insufficient phonetic similarity

One potential explanation for the failure to observe morphological priming is that the prime and target were not phonetically similar enough; i.e., maybe long-lag morphological priming only occurs when the prime and target are sufficiently similar in form (see also 1.2.2). This seems unlikely, though, given that long-lag morphological priming has consistently been observed in other morphological relationships that involve changes to the form of the stem, such that the exact form of the prime is not present in the target and vice versa – for instance, *sleep-slept* (Downie et al., 1985; Fowler et al., 1985), French [kuzin] ‘female cousin’ – [kuzɛ̃] ‘male cousin’ (Kouider & Dupoux, 2009), or Welsh [ben] – [pen] ‘head’ (Boyce et al., 1987). Morphological priming for primes and targets with substantially different forms is also common in masked priming experiments on Semitic roots, e.g., *tkabbar* ‘to be enlarged’ – *kiber* ‘to grow’ in Ussishkin et al. (2015). It is not obvious that the form difference between, e.g., *shi<sup>R</sup>* and *shi<sup>L</sup>* is larger than that between *sleep* and *slept*, [kuzin] and [kuzɛ̃] ‘female/male cousin’, etc.

Note, however, that in all the studies mentioned above, either the use of the different allomorph does not require a particular phonological context (e.g., an English speaker’s choice to produce *slept* instead of *sleep* is not dependent on phonological context) or the required phonological context is present in the experiment (as in Boyce et al., 1987); these situations are both different than that of the present study, in which the use of a Low- or Rising-tone variant depends on phonological context but that context was not provided within the experiment. This difference between the present study and previous study merits some extended consideration, below.

### 5.2.2 Lack of context

As mentioned above, the lack of morphological priming may be due to the lack of context to license the phonological alternation. Recall that a Mandarin Low-tone syllable changes to Rising tone if it’s followed by another Low-tone syllable within the same intonational unit; for example, *mai<sup>L</sup>* (“buy”) is Low-tone but it is pronounced with Rising tone (or something similar to Rising tone) in *mai<sup>R</sup> jiu<sup>L</sup>* (“buy wine”). Thus, in the present study when isolated morphemes were presented out of context, there may have been no reason for a participant hearing *mai<sup>L</sup>* to activate its Rising-tone variant, or for a participant hearing *mai<sup>R</sup>* to recognize it as a Rising-tone variant of *mai<sup>L</sup>*. This makes the present study very different than all other morphological priming studies we are aware of, which have presented morphemes within some sort of context that makes their canonical form recoverable. For most studies this involves morphemes within morphologically complex inflected or derived words like *assumption* (*assume* + *tion*) or *slept* (*sleep* plus a past tense morpheme), or morphemes whose own phonology inherently licenses the alternation (e.g., *Adam* and *atom*, which both license flapping by virtue of having an intervocalic alveolar stop after a stressed vowel). Morphological priming has also been shown, however, with morphemes



presented in a *syntactic* context that licenses a different surface form of that morpheme. For instance, Boyce and colleagues (1987) found morphological priming between different variants of Welsh words affected by consonant mutation. Many Welsh word-initial consonants undergo phonological alternation triggered by the lexical or syntactic context in which they occur; for example, /pen/ ('head') becomes [ben] in the phrase [ei ben o] ('his head'). This is arguably more similar to tone sandhi than the *assume-assumption* or *sleep-slept* sort of cases are, as both the Welsh consonant mutation and Mandarin tone sandhi cases involve a phonological change that can be triggered by word-external syntactic context. Crucially, however, Boyce and colleagues presented the words in phrasal contexts that licensed the consonant mutation (e.g., [ei ben o]). Thus, we cannot rule out the possibility that our study's failure to observe morphological priming is due to our study's lack of phonological contexts to license tone sandhi.

There is some precedent that supports the explanation based on context presented above. For instance, Gaskell and Marslen-Wilson (1996) found that cross-modal priming between phonological variants in a sentence context was larger when the prime was phonologically licensed by the context than when it was not. Indeed, the large literature on compensation for phonological alternations in context (as opposed to the priming literature based on words in isolation) provides substantial evidence that context matters and that variants often don't contact possible lexical entries unless a context allows it – for example, an auditory input 'greem' does not contact the lexical entry for *green* unless presented in a context such as 'greem boat', where 'greem' is a plausible assimilated form of *green* (for reviews see, e.g., Marslen-Wilson et al., 1995; Gaskell & Snoeren, 2008; Tavabi et al., 2009; Sun et al., 2015).

Another type of context that may moderate morphological priming is overall experimental context. A recent long-lag priming study by Gaston and colleagues (2021) suggests that previously-reported long-lag morphological priming effects may have been dependent on the presence of repeated morphological contexts over the course of an experiment. For instance, stem priming occurred for Kouider & Dupoux, 2009, when targets were always masculine forms of the feminine primes (*cousine* 'female cousin' – *cousin* 'male cousin', etc.), whereas stem priming was not robust for Gaston and colleagues when the morphological relationships were not repeated across items (i.e., one item was *hero-heroism*, one was *work-worker*, etc.).

It is unclear, however, whether these contextual factors can explain the findings of the present study. It could be argued that the targets and primes in the present study shared a close contextual relationship as in Kouider & Dupoux, 2009 (because primes and targets were always related through the same phonological rule), or that they did not (because they were presented out of context and thus the connection between a given stimulus and its phonological variants was not made explicit); thus, it's not clear whether Gaston and colleagues' (2021) account would predict the present manipulation to elicit or not elicit morphological priming. Furthermore, regarding the issue of phonological licensing contexts, a recent study testing a similar manipulation as the



present study but with immediate priming (Politzer-Ahles et al., 2022) failed to observe priming for phonological variants of the same Mandarin morpheme, regardless of whether these were or were not presented in a context that licenses the phonological change. This casts some doubt upon whether the lack of phonological context explains the failure to observe morphological priming in the present study. Nevertheless, further study on this is needed; it certainly is the case that the present study with morphemes presented out of context lacks analogues in the existing literature (we are unaware of any long-lag priming studies that used stimuli consisting of morphemes stripped away from their full-word context, e.g., *-ism* or *knive-*, or that tested questions like whether *hand* primes *ham* just because it can be pronounced this way in *handbag*), and thus investigating the priming of Mandarin phonological variants in a more traditional design – putting them either in a compound word context or a short phrasal context to license the tone sandhi – would be valuable.

In any case, it is clear that the nature of the relationship tested in the present study is quite different from, and thus not very comparable to, that tested in most previous studies of morphological long-lag priming. Whereas most studies of morphological priming have examined priming between two words which share a morpheme such that at least one word might be processed by decomposing it into its parts (e.g., *assume-assumption*, French *cousine-cousin*, *humanism-heroism*, etc.), the present study tested for priming between two different phonological realizations of the same morpheme, without any context (as if we had presented participants with the [s] and [z] realizations of English plural -s, but in isolation) and with no need for decomposition. Thus, the stimuli were divorced from any actual word formation processes, which makes it debateable whether the present manipulation could be considered “morphological” priming at all. Furthermore, many previous studies of morphological or allophonic priming examined situations in which the listener must obligatorily map the input onto a different lexical representation in order to identify it (for example, if flaps and unreleased stops are not canonical forms of any English phoneme, then hearing an input with a flap or unreleased stop (e.g., McLennan et al., 2003, 2005; Sumner & Samuel, 2005) may obligate the hearer to activate an underlying form that is different from the physical input), whereas the present study examined a case where activation of variants is not strictly necessary, because the input is already a possible realization of an existing canonical form, and thus a listener hearing a syllable with a Rising tone may have no reason to attempt to activate Low-tone forms, or vice versa. This raises the question of whether the presence of existing canonical forms pre-empts a process that would otherwise trigger activation of other phonological variants (see also McLennan et al.’s (2003, 2005) claims about ambiguity, discussed in section 5.1).

### 5.2.3 Homophony

Another way in which the present study is different from previous studies of long-lag morphological priming is homophony. Because Mandarin has a fairly limited syllable inventory and most morphemes are single-syllable, most morphemes (and most of the stimuli in the present

study) are homophonous: for example, *shi*<sup>L</sup> is the pronunciation of multiple morphemes, such as 始 ('start'), 史 ('history'), 矢 ('arrow'), 驶 ('drive'), 使 ('envoy'), or 屎 ('shit'). This is not the case for previous morphological priming studies, in which the morphemes within the prime and target (e.g., *assume* and *-tion* in *assumption*, or *cousin* plus a feminine morpheme in French *cousine*) can be uniquely identified. Is it possible, then, that morphological priming depends on being able to uniquely identify and activate a particular morpheme, and that that activation does not occur (and thus does not lead to priming) if the stimulus is not uniquely associated with a particular morpheme? This could be tested in a design similar to that of the present study by limiting the stimuli to morphemes with no homophones (although the available set of such stimuli in Mandarin would be very small) or by seeing if the size of the morphological priming effect is correlated with the number of (or just the presence vs. absence of) homophones. One aspect of our results, however, already speaks against this account: robust identity priming was observed with the same stimuli that failed to elicit morphological priming, and this identity priming could not be explained as a mere episodic effect. If priming depends on uniquely identifying a particular morpheme, then it is not clear how genuinely lexical identity priming could have occurred in the present study. Another piece of evidence against the argument that morphological priming depends on unique identification of one morpheme is that McLennan and colleagues (2003) find priming with ambiguous primes (e.g., [æɾəm], which is ambiguous between *atom* and *Adam*, can prime both), although their stimuli are less ambiguous than (i.e., map onto fewer possible underlying forms than) ours.

To think about the potential effects of homophony on the present results, it is important to keep in mind that long-lag priming – even “morphological” long-lag priming – is often understood as relying on the activation of orthographic or phonological codes rather than the activation of morphemes themselves (e.g., Bowers, 2000; Bowers & Kouider, 2003). Long-lag priming even occurs for pseudowords, which have no morphological entries associated with them. Therefore, it is not clear that obtaining priming in the present study should rely on the participant’s being able to uniquely associate each stimulus to one morpheme.

Thinking about long-lag priming in terms of access codes offers another way to contrast the present study with other studies on priming across variants. In the present study, a form like *shi*<sup>R</sup> is not only an different phonological code for the morpheme *shi*<sup>L</sup>; it also happens to be the phonological code for completely different morphemes, whose canonical pronunciations are *shi*<sup>R</sup>. This is different from the sort of episodic variation examined in many previous priming studies; for example, when the same word is spoken by different speakers of the same dialect (e.g., Luce & Lyons, 1998), it is not the case that the second token matches a completely different morpheme than the first token did. Recall that this sort of ambiguity was also implicated in the explanation of previous long-lag priming results (see the discussion of McLennan et al., 2003, 2005, in 5.1). A valuable avenue for future research could be to systematically distinguish between priming

observed in cases where the phonological/orthographic code activated by the target does not match any morpheme other than that of the prime (this is the case in typical studies of episodic variation, e.g., Luce & Lyons, 1998); cases where the code activated by the target is an equally good match for this morpheme and some other morpheme (e.g., McLennan and colleagues, 2003, 2005, where for instance a target form [ɹeɪɹə] is an equally good match for *raider* and *rater*, one of which was used as a prime and one of which was not); and cases where the code activated by a target is a poor match for the primed code and a better match for something else (the present study).

### 5.2.4 Competition

Finally, another possible explanation for the present study's failure to observe morphological priming might be that priming did indeed occur but was cancelled out by competition. Other studies have also suggested that competition between allomorphs, and ensuing suppression of the activation of one allomorph, may cancel out the priming benefit that would otherwise obtain from repeated activation of the same root morpheme. For instance, Stockall and Marantz (2006) found no facilitation for present-tense targets by formally similar past-tense roots in an immediate paired priming experiment. They speculate that this may be because these forms compete for recognition and thus the recognition of the past-tense prime inhibits the present-tense target. While that was an immediate priming experiment, Stockall (2004) found similar results in long-lag priming, and they could be explained in the same way.

It is also possible that previous activation of a morpheme facilitates later re-activation of the same morpheme but makes lexical decision to different forms of this morpheme more difficult – for example, hearing a prime *shi<sup>L</sup>* might activate morphemes associated with this syllable, and later on it might be relatively easy to re-activate these morphemes upon hearing a target *shi<sup>R</sup>*, but when the participant needs to make a conscious decision about whether or not *shi<sup>R</sup>* is a word then this decision will become harder when they are distracted by the heightened activation of morphemes whose canonical pronunciation is *shi<sup>L</sup>*. If this explanation is right, then facilitation might be observed with a different task, or it might be possible to see a pattern of early facilitation followed by later inhibition when using a measure that allows for a more fine-grained probing of the time course of cognition, such as event-related brain potentials; such a pattern might also be consistent with the abovementioned results of Stockall (2004) and Stockall and Marantz (2006), who found facilitation (of past-tense targets by present-tense primes) in immediate priming but not in long-lag priming. If these sets of results are to be taken as evidence for competition between allomorphs, a task remaining for future study is to delineate the circumstances under which competition occurs and is sufficient to cancel out priming, versus the circumstances under which morphological priming does obtain either because competition does not occur or because the competition is not sufficient to overcome the facilitation due to priming.

### 5.3 A directional asymmetry between Low and Rising tones

Not only did morphologically-related primes unexpectedly inhibit activation of targets across all three experiments, but this pattern was driven by the Rising targets (preceded by Low primes) in all three experiments. Specifically, in Experiments 1, 2, and 3, Rising targets preceded by Low primes were responded to 33, 30, and 16 ms more slowly than Rising targets preceded by unrelated primes. On the other hand, Low targets preceded by Rising primes were responded to 9, -5, and -2 ms more slowly than the same targets preceded by unrelated primes in Experiments 1, 2, and 3, respectively. (Note that the manipulation of the target's tone was within items in Experiment 1, but between items in Experiments 2 and 3.)

Identity priming did not show an asymmetry like this; the identity priming effects in Experiments 2 and 3 were 59 and 56 ms for Rising targets, and 60 and 63 ms for Low targets. The directional asymmetry thus seems to be limited to how Low and Rising forms activate their variant forms, not their canonical forms. There was, however, a comparable asymmetry in form-only conditions: across experiments, Rising targets were inhibited by 13 ms if they were preceded by High- or Falling-tone primes (compared to completely unrelated primes), whereas Low tones were primed/inhibited by 0 ms when preceded by High- or Falling-tone primes. Unfortunately our critical stimuli did not include form-related priming conditions allowing us to compare the effects of Low and Rising primes,<sup>5</sup> which means that, while we can claim that Low and Rising targets were subject to different amounts of form priming, we cannot be sure whether Low and Rising *primes caused* different amounts of form priming.

We did not expect such asymmetries, and are not aware of previous priming studies on allophonic variants reporting comparable asymmetries. There is, however, some precedent for Low-Rising tone asymmetries outside of the priming literature, specifically in electrophysiological studies examining habituation effects in an oddball paradigm. Li and Chen (2015) and Politzer-Ahles and colleagues (2016) both found that an unexpected Low tone after a series of Rising tones is treated as a bigger mismatch than an unexpected Rising tone after a series of Low tones. We are not yet aware of any satisfactory explanation for this pattern. Li and Chen (2015) argue that this pattern happens because Low-tone morphemes have multiple stored variants (one which is Low and one which is similar to a Rising tone), and hearing Low tone activates both, so that the unexpected Rising tone later does not strongly mismatch with the Rising-like variant that has already been activated. This can successfully explain their own results but does not on its own explain the results of Politzer-Ahles et al. (2016), who found that series of Low tones also were not strongly mismatched by other unexpected tones either. Politzer-Ahles et al. (2016) argued that Low tone may involve a phonologically underspecified feature, but this is just a claim they

---

<sup>5</sup> Our critical items all used Low or Rising targets and the form-related priming conditions were created by using High or Falling primes; we had form-related priming with High/Falling targets and Low/Rising primes in fillers, but these were not designed to allow a comparison between Low and Rising primes, so only a between-items comparison of Low and Rising primes in the fillers is possible.

put forth to account for their results and there is no independent phonological evidence for it (see also Meng et al., 2021, for criticism of the argument that Low tone is underspecified). Meng et al. (2021) suggest that this asymmetry may happen because a series of multiple Low tones is phonologically illegal, but this is probably not the case for the habituation paradigms in these studies, where the stimuli are presented with pauses in between (400 ms in Li and Chen, 2015, and 500 ms in Politzer-Ahles et al., 2016), which means that they are presumably not in the same intonational phrase and thus not in a context that would license tone sandhi. Given that there is not yet any well-supported explanation for what mechanism causes these asymmetries in electrophysiological studies, it is also not yet clear whether the asymmetry in the current study is caused by the same mechanism or something else.

Might ambiguity play some role in this asymmetrical effect? Rising-tone syllables are potentially ambiguous, because they could be variants of underlyingly Rising or underlyingly Low tones. On the other hand, Low-tone syllables are unambiguous, because only underlyingly Low syllables can surface as Low. Crucially, Low (unambiguous) primes pretty reliably inhibited Rising targets in our experiments, whereas Rising (ambiguous) primes had much less inhibition on Low targets.

It would be tempting to explain this pattern by positing that primes activate all underlying forms that they might be surface forms of, and cause inhibition for any similar forms which are not activated (or, as discussed in the previous section, interfere with the lexical decision task carried out on similar forms). Under this explanation, Low primes activate Low representations and suppress Rising representations; Rising primes, on the other hand, activate both. This could explain why we observed inhibition for Rising targets after Low primes (which, under this account, do not activate Rising variants) but not for Low targets after Rising primes (which, under this account, do activate both Low and Rising variants). There are, however, several problems with this account. Most seriously, the difference between our identity priming results and morphological priming results is not predicted by this account: Rising primes facilitated Rising targets in identity priming, but they didn't facilitate Low targets in morphological priming, they merely failed to inhibit them. Some additional mechanism would be needed to explain why Rising primes activate Low representations enough to keep there from being inhibition in the Rising-Low priming condition, but not enough for there to actually be facilitation. Secondly, this explanation relies on the assumption that form-only relationships cause inhibition in long-lag priming, which has generally not been observed in extant research.<sup>6</sup> Finally, this explanation also

---

<sup>6</sup> This may be explained by assuming that form relationships based on tone cause different effects in long-lag priming than the segmental or orthographic form relationships tested in most previous long-lag priming research; the "form" relationships in the present study were based on a one-tone difference between prime and target, and there is a substantial body of research suggesting that tone does not contribute to lexical access in the same way segments do (e.g., Sereno & Lee, 2015; Wiener & Turnbull, 2015; see also Chen et al., 2002, and O'Seaghda et al., 2010, for similar evidence regarding the role of tone and segments in speech encoding).

cannot explain why a similar asymmetry emerged for Low and Rising targets preceded by purely form-related High and Falling primes in Experiment 1, where Rising targets had a 12 ms form inhibition effect whereas Low targets only had a 3 ms form inhibition effect. That suggests that the explanation for the asymmetry should be based on the processing of the target, rather than the processing of the prime, since the targets are what is in common across all these comparisons.

Overall, while this Low-Rising asymmetry is tantalizing and its robustness across experiments suggests that there is something important causing it, we cannot yet offer any satisfactory explanation for it and can only conclude that future research is needed to investigate what causes this asymmetry.

## **6. Conclusion**

We tested whether exposure to one form of a morpheme facilitates responses to another form of the same morpheme encountered later, by having Mandarin-speaking listeners make lexical decisions to syllables that represent possible pronunciations of morphemes. The answer to the question was unambiguously no: across three experiments we failed to find priming between different pronunciations of the same morpheme, even though identity priming (priming between the same syllable) was robust. Overall, although there are many differences between the present study and previous studies of morphological or allophonic long-lag priming, the failure to elicit priming in this study hints that there may be important boundary conditions on when morphemes are activated – although the details of what these boundary conditions are remain to be elucidated in further study. While the present study offers more questions than answers, it offers a hopefully useful observation about the limits of the extent to which morpheme recognition adjusts for phonological alternation.

---

## Data accessibility statement

All materials, data, and analysis code associated with this study are available on OSF at <https://osf.io/pt8qv/>.

## Ethics and consent

Procedures for the experiment were approved by the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University (reference: HSEARS20160918002)

## Acknowledgements

This work was supported by grant #G-YBZM from the Department of Chinese and Bilingual Studies at The Hong Kong Polytechnic University to SPA.

## Competing interests

The authors have no competing interests to declare.

## Author contributions

**Stephen Politzer-Ahles:** Conceptualization, methodology, formal analysis, writing – original draft, writing – review & editing, visualization, supervision, funding acquisition

**Lei Pan:** methodology, investigation, writing – review & editing

**Jueyao Lin:** methodology, investigation, writing – review & editing

**Ka Keung Lee:** investigation, writing – review & editing

---

## References

- Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. DOI: <https://doi.org/10.1016/j.jml.2007.12.005>
- Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1059–1082. DOI: <https://doi.org/10.1037/0278-7393.23.5.1059>



- Bentin, S., & Feldman, L. (1990). The contribution of morphological and semantic relatedness to repetition priming at short and long lags: Evidence from Hebrew. *The Quarterly Journal of Experimental Psychology*, 42, 693–711. DOI: <https://doi.org/10.1080/14640749008401245>
- Bowers, J. (1996). Different perceptual codes support priming for words and pseudowords: Was Morton right all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1336–1353. DOI: <https://doi.org/10.1037//0278-7393.22.6.1336>
- Bowers, J. (2000). In defense of abstractionist theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, 7, 83–99. DOI: <https://doi.org/10.3758/BF03210726>
- Bowers, J., & Kouider, S. (2003). Developing theories of priming with an eye on function. In J. S. Bowers & C. Marsolek (Eds.), *Rethinking Implicit Memory*, pp. 19–40. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780192632326.003.0002>
- Bowers, J., & Michita, Y. (1998). An investigation into the structure and acquisition of orthographic knowledge: Evidence from cross-script kanji-hiragana priming. *Psychonomic Bulletin & Review*, 5, 259–264. DOI: <https://doi.org/10.3758/BF03212948>
- Boyce, S., Browman, D., & Goldstein, L. (1987). Lexical organization and Welsh consonant mutation. *Journal of Memory and Language*, 26, 419–452. DOI: [https://doi.org/10.1016/0749-596X\(87\)90100-8](https://doi.org/10.1016/0749-596X(87)90100-8)
- Brown, H., Sharma, N., & Kirsner, K. (1984). The role of script and phonology in lexical representation. *The Quarterly Journal of Experimental Psychology*, 36A, 491–505. DOI: <https://doi.org/10.1080/14640748408402173>
- Burt, J., & Long, J. (2011). Are word representations abstract or instance-based? Effects of spelling inconsistency in orthographic learning. *Canadian Journal of Experimental Psychology*, 65, 214–228. DOI: <https://doi.org/10.1037/a0023708>
- Chandrasekaran, B., Krishnan, A., & Gandour, J. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, 1128, 148–156. DOI: <https://doi.org/10.1016/j.brainres.2006.10.064>
- Chen, J., & Chen, T. (in press). Mechanism and locus of tone 3 sandhi during Mandarin Chinese spoken word production. *Journal of Chinese Linguistics Monograph Series*.
- Chen, J., Chen, T., & Dell, G. (2002). Wordform encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, 46, 751–781. DOI: <https://doi.org/10.1006/jmla.2001.2825>
- Chien, Y., Sereno, J., & Zhang, J. (2016). Priming the representation of Mandarin tone 3 sandhi words. *Language, Cognition, and Neuroscience*, 31, 179–189. DOI: <https://doi.org/10.1080/23273798.2015.1064976>
- Chuang, C., & Hiki, S. (1972). Acoustical features and perceptual cues of the four tones of standard colloquial Chinese. *The Journal of the Acoustical Society of America*, 52, 146. DOI: <https://doi.org/10.1121/1.1981919>
- Coane, H., & Balota, D. (2010). Repetition priming across distinct contexts: Effects of lexical status, word frequency, and retrieval test. *The Quarterly Journal of Experimental Psychology*, 63, 2376–2398. DOI: <https://doi.org/10.1080/17470211003687546>

- Connine, C., & Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review*, 23, 235–245. DOI: <https://doi.org/10.1515/TLR.2006.009>
- Creemers, A., Goodwin Davies, A., Wilder, R., Tamminga, M., & Embick, D. (2020). Opacity, transparency, and morphological priming: A study of prefixed verbs in Dutch. *Journal of Memory and Language*, 110, 104055. DOI: <https://doi.org/10.1016/j.jml.2019.104055>
- Dannenbring, G., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. *Canadian Journal of Psychology*, 36, 435–444. DOI: <https://doi.org/10.1037/h0080650>
- Davelaar, E., & Coltheart, M. (1975). Effects of interpolated items on the association effect in lexical decision tasks. *Bulletin of the Psychonomic Society*, 6, 269–272. DOI: <https://doi.org/10.3758/BF03336658>
- De Grauwe, S., Lemhöfer, K., Willems, R., & Schriefers, H. (2014). L2 speakers decompose morphologically complex verbs: fMRI evidence from priming of transparent derived verbs. *Frontiers in Human Neuroscience*, 8, 802. DOI: <https://doi.org/10.3389/fnhum.2014.00802>
- Dohmes, P., Zwitserlood, P., & Bölte, J. (2004). The impact of semantic transparency of morphologically complex words on picture naming. *Brain and Language*, 90, 203–212. DOI: [https://doi.org/10.1016/S0093-934X\(03\)00433-4](https://doi.org/10.1016/S0093-934X(03)00433-4)
- Downie, R., Milech, D., & Kirsner, K. (1985). Unit definition in the mental lexicon. *Australian Journal of Psychology*, 37, 141–155. DOI: <https://doi.org/10.1080/00049538508256394>
- Drews, E., & Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1098–1116. DOI: <https://doi.org/10.1037/0096-1523.21.5.1098>
- Durgunoğlu, A., & Roediger, H. (1987). Test differences in accessing bilingual memory. *Journal of Memory and Language*, 26, 377–391. DOI: [https://doi.org/10.1016/0749-596X\(87\)90097-0](https://doi.org/10.1016/0749-596X(87)90097-0)
- Emmorey, K. (1991). Repetition priming with aspect and agreement morphology in American Sign Language. *Journal of Psycholinguistic Research*, 20, 365–388. DOI: <https://doi.org/10.1007/BF01067970>
- Feldman, L. (1992). Morphological relationships revealed through the repetition priming task. In M. Noonan, P. Downing & S. Lima (Eds.), *Linguistics & literacy* (pp. 239–254). Amsterdam/Philadelphia: Benjamins. DOI: <https://doi.org/10.1075/tsl.21.18fel>
- Feldman, L. (2000). Are morphological effects distinguishable from the effects of shared meaning and shared form? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1431–1444. DOI: <https://doi.org/10.1037/0278-7393.26.6.1431>
- Feldman, L., & Fowler, C. (1987). The inflected noun system in Serbo-Croatian: Lexical representation of morphological structure. *Memory and Cognition*, 15, 1–12. DOI: <https://doi.org/10.3758/BF03197707>
- Feldman, L., & Moskovljević, J. (1987). Repetition priming is not purely episodic in origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 573–581. DOI: <https://doi.org/10.1037/0278-7393.13.4.573>

- Feldman, L., & Siok, W. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language*, *40*, 559–576. DOI: <https://doi.org/10.1006/jmla.1998.2629>
- Forster, K., & Forster, J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124. DOI: <https://doi.org/10.3758/BF03195503>
- Fowler, C., Napps, S., & Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory and Cognition*, *13*, 241–255. DOI: <https://doi.org/10.3758/BF03197687>
- Gaskell, G., & Marslen-Wilson, W. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 144–158. DOI: <https://doi.org/10.1037/0096-1523.22.1.144>
- Gaskell, M., & Snoeren, N. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1632–1647. DOI: <https://doi.org/10.1037/a0011977>
- Gaston, P., Stockall, L., VanWagenen, S., & Marantz, A. (2021). Memory for affixes in a long-lag priming paradigm. *Glossa: A Journal of General Linguistics*, *6*. DOI: <https://doi.org/10.16995/glossa.5735>
- Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279. DOI: <https://doi.org/10.1037/0033-295X.105.2.251>
- Gong, S., & Zhang, J. (2021). Modelling Mandarin speakers' phonotactic knowledge. *Phonology*, *38*, 241–275. DOI: <https://doi.org/10.1017/S0952675721000166>
- Hallé, P., Chéreau, C., & Segui, J. (2000). Where is the /b/ in “absurde” [apsyrd]? It is in French listeners' minds. *Journal of Memory and Language*, *43*, 618–639. DOI: <https://doi.org/10.1006/jmla.2000.2718>
- Hanson, V., & Wilkenfeld, D. (1985). Morphophonology and lexical organization in deaf readers. *Language and Speech*, *28*, 269–80. DOI: <https://doi.org/10.1177/002383098502800304>
- Hay, J., & Baayen, R. (2005). Shifting paradigms: Gradient structure in morphology. *TRENDS in Cognitive Sciences*, *9*, 342–348. DOI: <https://doi.org/10.1016/j.tics.2005.04.002>
- Horner, A., & Henson, R. (2011) Stimulus–response bindings code both abstract and specific representations of stimuli: Evidence from a classification priming design that reverses multiple levels of response representation. *Memory and Cognition*, *39*, 1457–1471. DOI: <https://doi.org/10.3758/s13421-011-0118-8>
- Kempler, S., & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology*, *73*, 441–454. DOI: <https://doi.org/10.1111/j.2044-8295.1982.tb01826.x>
- Kirsner, K., Smith, M., Lockhart, R., King, M., & Jain, M. (1984). The bilingual lexicon: Language-specific units in an integrated network. *Journal of Verbal Learning and Verbal Behavior*, *23*, 519–539. DOI: [https://doi.org/10.1016/S0022-5371\(84\)90336-0](https://doi.org/10.1016/S0022-5371(84)90336-0)

- Koester, D., & Schiller, N. (2008). Morphological priming in overt language production: Electrophysiological evidence from Dutch. *NeuroImage*, *42*, 1622–1630. DOI: <https://doi.org/10.1016/j.neuroimage.2008.06.043>
- Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science*, *16*, 617–625. DOI: <https://doi.org/10.1111/j.1467-9280.2005.01584.x>
- Kouider, S., & Dupoux, E. (2009). Episodic accessibility and morphological processing: Evidence from long-term auditory priming. *Acta Psychologica*, *130*, 38–47. DOI: <https://doi.org/10.1016/j.actpsy.2008.09.011>
- Lensink, S., Verdonschot, R., & Schiller, N. (2014). Morphological priming during language switching: An ERP study. *Frontiers in Human Neuroscience*, *8*, 995. DOI: <https://doi.org/10.3389/fnhum.2014.00995>
- Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PloS ONE*, *10*, e0143097. DOI: <https://doi.org/10.1371/journal.pone.0143097>
- Luce, P., & Lyons, E. (1998). Specificity for memory representations of spoken words. *Memory and Cognition*, *26*, 708–715. DOI: <https://doi.org/10.3758/BF03211391>
- Luce, P., McLennan, C., & Charles-Luce, J. (2003). Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory*. Oxford, England: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780192632326.003.0009>
- Marslen-Wilson, W., Nix, A., & Gaskell, G. (1995). Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes*, *10*, 285–308. DOI: <https://doi.org/10.1080/01690969508407097>
- Marslen-Wilson, W., & Tyler, L. (1998). Rules, representations, and the English past tense. *Trends in Cognitive Sciences*, *2*, 428–435. DOI: [https://doi.org/10.1016/S1364-6613\(98\)01239-X](https://doi.org/10.1016/S1364-6613(98)01239-X)
- McLennan, C., & Luce, P. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 306–321. DOI: <https://doi.org/10.1037/0278-7393.31.2.306>
- McLennan, C., Luce, P., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 539–553. DOI: <https://doi.org/10.1037/0278-7393.29.4.539>
- McLennan, C., Luce, P., & Charles-Luce, J. (2005). Representation of lexical form: Evidence from studies of sublexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1308–1314. DOI: <https://doi.org/10.1037/0096-1523.31.6.1308>
- Meng, Y., Wynne, H., & Lahiri, A. (2021). Representation of “T3 sandhi” in Mandarin: Significance of context. *Language, Cognition and Neuroscience*, *36*, 791–808. DOI: <https://doi.org/10.1080/2373798.2021.1893769>
- Münte, T., Say, T., Clahssen, H., Schiltz, K., & Kutas, M. (1999). Decomposition of morphologically complex words in English: Evidence from event-related potentials. *Cognitive Brain Research*, *7*, 241–253. DOI: [https://doi.org/10.1016/S0926-6410\(98\)00028-7](https://doi.org/10.1016/S0926-6410(98)00028-7)

- Murrell, G., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology*, *102*, 963–968. DOI: <https://doi.org/10.1037/h0036551>
- Napps, S. (1989). Morphemic relationships in the lexicon: are they distinct from semantic and formal relationships? *Memory & Cognition*, *17*, 729–739. DOI: <https://doi.org/10.3758/BF03202634>
- Napps, S., & Fowler, C. (1987). Formal relationships among words and the organization of the mental lexicon. *Journal of Psycholinguistic Research*, *16*, 257–271. DOI: <https://doi.org/10.1007/BF01067546>
- Nixon, J., Chen, Y., & Schiller, N. (2014). Multi-level processing of phonetic variants in speech production and visual word processing: Evidence from Mandarin lexical tones. *Language, Cognition, and Neuroscience*, *30*, 1–15. DOI: <https://doi.org/10.1080/23273798.2014.942326>
- O’Séaghdha, P., Chen, J., & Chen, T. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, *115*, 282–302. DOI: <https://doi.org/10.1016/j.cognition.2010.01.001>
- Pierrehumbert, J. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, *2*, 33–52. DOI: <https://doi.org/10.1146/annurev-linguist-030514-125050>
- Politzer-Ahles, S., Lin, J., Pan, L., & Lee, K. (2022). N400 evidence that the early stages of lexical access ignore knowledge about phonological alternations. *Language and Speech*, *65*, 354–376. DOI: <https://doi.org/10.1177/00238309211020026>
- Politzer-Ahles, S., & Piccinini, P. (2018). On visualizing phonetic data from repeated measures experiments with multiple random effects. *Journal of Phonetics*, *70*, 56–69. DOI: <https://doi.org/10.1016/j.wocn.2018.05.002>
- Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 1547–1570. DOI: <https://doi.org/10.1037/xhp0000242>
- Pulvermüller, F. (1999). Words in the brain’s language. *Behavioral and Brain Sciences*, *22*, 253–279. DOI: <https://doi.org/10.1017/S0140525X9900182X>
- R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Ranbom, L. & Connine, C. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*, 273–298. DOI: <https://doi.org/10.1016/j.jml.2007.04.001>
- Roediger, H., & Challis, B. (1992). Effects of exact repetition and conceptual repetition on free recall and primed word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 3–14. DOI: <https://doi.org/10.1037/0278-7393.18.1.3>
- Rueckl, J., & Aicher, K. (2008). Are CORNER and BROTHER morphologically complex? Not in the long term. *Language and Cognitive Processes*, *7–8*, 972–1001. DOI: <https://doi.org/10.1080/01690960802211027>



- Sciama, S., & Dowker, A. (2007). Abstraction and perceptual individuation in primed word identification are modulated by distortion and repetition: A dissociation. *Memory, 15*, 899–911. DOI: <https://doi.org/10.1080/09658210701757305>
- Sereno, J., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech, 58*, 131–151. DOI: <https://doi.org/10.1177/0023830914522956>
- Sharma, B., Liu, C., & Yao, Y. (2015). Perceptual confusability of Mandarin sounds, tones and syllables. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0890. Retrieved from <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0890.pdf>.
- Stanners, R., Neiser, J., Herson, W., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior, 18*, 399–412. DOI: [https://doi.org/10.1016/S0022-5371\(79\)90219-6](https://doi.org/10.1016/S0022-5371(79)90219-6)
- Stockall, L. (2004). *Magnetoencephalographic investigations of morphological identity and irregularity*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Stockall, L., & Marantz, A. (2006). A single-route, full decomposition model of morphological complexity. *The Mental Lexicon, 1*, 85–123. DOI: <https://doi.org/10.1075/ml.1.1.07sto>
- Stoltz, J., & Feldman, L. (1995). The role of orthographic and semantic transparency of the base morpheme in morphological processing. In L. Feldman (Ed.), *Morphological aspects of language processing*. Hillsdale: Erlbaum.
- Sumner, M. (2013). A phonetic explanation of phonological variant effects. *Journal of the Acoustical Society of America, 134*, EL26–EL32. DOI: <https://doi.org/10.1121/1.4807432>
- Sumner, M., & Samuel, A. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language, 52*, 322–338. DOI: <https://doi.org/10.1016/j.jml.2004.11.004>
- Sun, Y., Giavazzi, M., Adda-Decker, M., Barbosa, L., Kouider, S., Bachoud-Lévi, A., Jacquemot, C., & Peperkamp, S. (2015). Complex linguistic rules modulate early auditory brain responses. *Brain and Language, 149*, 55–65. DOI: <https://doi.org/10.1016/j.bandl.2015.06.009>
- Tavabi, K., Elling, L., Dobel, C., Pantev, C., & Zwitserlood, P. (2009). Effects of place of articulation changes on auditory neural activity: A magnetoencephalography study. *PLoS ONE, 4*, e4452. DOI: <https://doi.org/10.1371/journal.pone.0004452>
- Tenpenny, P. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 2*, 339–363. DOI: <https://doi.org/10.3758/BF03210972>
- Tsang, Y., Wong, A., Huang, J., & Chen, H. (2014). Morpho-orthographic and morpho-semantic processing in word recognition and production: Evidence from ambiguous morphemes. *Language, Cognition and Neuroscience, 29*, 543–560. DOI: <https://doi.org/10.1080/01690965.2013.790554>

- Ussishkin, A., Dawson, C., Wedel, A., & Schluter, K. (2015). Auditory masked priming in Maltese spoken word recognition. *Language, Cognition and Neuroscience*, *30*, 1096–1115. DOI: <https://doi.org/10.1080/23273798.2015.1005635>
- VanWagenen, S. (2014). The morphologically organized mental lexicon: Further experimental evidence. In C. T. Schütze & L. Stockall (Eds.), *Connectedness: Papers by and for Sarah VanWagenen, UCLA Working Papers in Linguistics*, *18* (pp. 1–29). <http://phonetics.linguistics.ucla.edu/wpl/issues/wpl18/papers/vanwagenen.pdf>
- Weyerts, H., Münte, T., Smid, H., & Heinze, H. (1996). Mental representations of morphologically complex words: An event-related potential study with adult humans. *Neuroscience Letters*, *206*, 125–128. DOI: [https://doi.org/10.1016/S0304-3940\(96\)12442-3](https://doi.org/10.1016/S0304-3940(96)12442-3)
- Wheeldon, L., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *The Quarterly Journal of Experimental Psychology*, *44A*, 723–761. DOI: <https://doi.org/10.1080/14640749208401307>
- Wiener, S., & Turnbull, T. (2015). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech*, *59*, 59–82. DOI: <https://doi.org/10.1177/0023830915578000>
- Wilder, R., Goodwin Davies, A., & Embick, D. (2019). Differences between morphological and repetition priming in auditory lexical decision: Implications for decompositional models. *Cortex*, *116*, 122–142. DOI: <https://doi.org/10.1016/j.cortex.2018.10.007>
- Zhang, J. (2010). Issues in the analysis of Chinese tone. *Language and Linguistics Compass*, *4*, 1137–1153. DOI: <https://doi.org/10.1111/j.1749-818X.2010.00259.x>
- Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes*, *10*, 545–600. DOI: <https://doi.org/10.1080/01690969508407114>
- Zuraw, K., Lin, I., Yang, M., & Peperkamp, S. (2021). Competition between whole-word and decomposed representations of English prefixed words. *Morphology*, *31*, 201–237. DOI: <https://doi.org/10.1007/s11525-020-09354-6>
- Zwitzerlood, P., Bölte, J., & Dohmes, P. (2000). Morphological effects on speech production: Evidence from picture naming. *Language and Cognitive Processes*, *15*, 563–591. DOI: <https://doi.org/10.1080/01690960050119706>

