Lexical tone in word activation

Yang, Q.

Chapter 1

General Introduction

In our minds, words are not organized in isolation but rather stored in interconnected networks, forming a complex web of associations (e.g., Collins & Loftus, 1975; McClelland & Rumelhart, 1985; Hutchison, 2003). Even when attempting to comprehend or produce a single word independently, we cannot simply match its meaning with its corresponding sound; instead, we involuntarily activate a list of related or similar words from our mental lexicon (Traxler, 2011).

For spoken word recognition, perhaps the most convincing evidence for parallel activation comes from the eye-tracking visual world paradigm (e.g., see Huettig et al., 2011 for a detailed review). In this paradigm (e.g., Allopenna et al., 1998), listeners are often asked to follow spoken instructions (e.g., "pick up the beaker") and identify the target (e.g., "beaker") from a few pictures displayed on the computer screen. Among the pictures, there is generally one phonological competitor which shares phonological overlaps with the target (e.g., target "beaker" – phonological competitor "beetle"), and two distractors that are unrelated to the target (e.g., target "beaker" – unrelated distractor "carriage"). During the process of identifying the correct target (e.g., "beaker"), listeners' eye movements are recorded, and they are found to fixate more on the phonological competitors (e.g., "beetle") than unrelated distractors (e.g., "carriage"). The fact that phonological competitors draw more visual attention than unrelated ones suggests that, while recognizing the target spoken word, listeners automatically activate phonologically related words and temporally consider them as potential word candidates for selection.

As for spoken word production, the most convincing evidence for parallel activation comes from the picture-word interference paradigm (Rosinski et al., 1975). In this paradigm, speakers are instructed to name pictures while ignoring the distractor word superimposed on the picture. Although phonologically related words appear to interfere with the process of spoken word recognition, they are found to facilitate the process of spoken word production (e.g., Meyer & Schriefers, 1991; Schriefers et al., 1990; Jescheniak & Schriefers, 2001). Specifically, if the distractor word is phonologically related to the target picture's

name (e.g., target "dog" – phonological distractor "fog"), speakers generally take less time to name the target picture compared with unrelated distractors (e.g., target "dog" – unrelated distractor "roof"). A generally accepted explanation of this effect is that the presence of the phonological distractor activated its corresponding sound in the speakers' mind, which overlaps with the target's sound form; speakers thus received not only top-down activation from the selected target lemma but also extra bottom-up activation from the phonological distractor during picture naming (e.g., Roelofs, 2000). Moreover, if the distractor word is categorically related to the target (e.g., target "dog" – categorically related semantic distractor "cat"), speakers take a longer time to name the target picture compared with unrelated distractors (e.g., Costa et al., 2003; La Heij, 1988; Lupker, 1984; Schriefers et al., 1990). A commonly suggested explanation for this effect is that the retrieval of the concept of the target picture not only activates the target word (e.g., "dog") but also words that are categorically related to the target (e.g., "cat"), which receive activation from both the target picture and the distractor word. Compared with unrelated distractors (e.g., "roof"), which only receive activation from the present distractor word itself, the activation level of the categorically related semantic distractor is higher and thus it is more demanding for speakers to select the intended target word (e.g., Levelt et al., 1999).

Both the processes of spoken word recognition and production are incremental, i.e., listeners and speakers do not just wait until the selection of a word to map meaning or plan for articulation. Rather, listeners and speakers activate multiple related words and their word forms for parallel processing. Besides evidence from the visual world paradigm and the picture-word interference paradigm, the co-activation of multiple related representations have also been validated by behavioural data of various tasks (e.g., lexical decision tasks, priming paradigms, blocked cyclic naming) and recent neuroimaging and electrophysiological data (see Nozari & Pinet, 2019 for a review). As one of the core principles of how our mind retrieves words, parallel activation has been

incorporated into most theories of language comprehension and production (e.g., Chen & Mirman, 2012; Dell, 1986; Levelt et al., 1999; McClelland & Elman, 1986).

During the last decades, research on parallel activation has also been extended to bilinguals, which provides a great opportunity to improve our understanding of the mental lexicon and lexical access. One critical issue that lies at the heart of the bilingual literature is whether bilingual lexical access is language-specific or language-nonspecific selection. Specifically, a central question is whether bilinguals co-activate words of the non-target language as well when comprehending or producing words in a target language. Classic paradigms introduced above, i.e., the visual world paradigm and the picture-word interference paradigm, have been applied to explore this issue. In a seminal eye-tracking study by Spivey and Marian (1999), Russian-English bilinguals were asked to follow instructions such as *Poloji marku nije krestika* "Put the stamp below the cross" and move objects on a whiteboard while their eye movements were being recorded. In critical trials, objects such as "marker" which share initial phonetic features with *marku* "stamp" were also presented. Eye movement analysis showed that the cross-language homophone "marker" attracted participants' visual attention from the target *marku* "stamp" significantly more than that of the unrelated control stimulus object (e.g., *lineika* "ruler"). This has been taken as evidence for parallel activation of words of two languages during spoken word recognition. As for spoken word production, evidence from the picture-word interference paradigm shows that phonologically and semantically related distractors of different languages manipulate the speed and accuracy of target picture naming (e.g., target "dog" – cross-language phonological distractor *muñeca* "doll" – cross-language semantic distractor *gato* "cat" - translation distractor "perro"), similar as distractors of the same language (e.g., target "dog" – phonological distractor "doll" – semantic distractor "cat"; see Hall, 2010 for a review). Such cross-language interaction provides strong evidence for the co-activation of bilinguals' two languages during spoken word production. There is

a consensus that bilinguals automatically access words of both their languages in speech comprehension and production (see Kroll et al., 2012 for a detailed review).

Although language non-specific parallel activation is widely agreed upon, it is important to note that most previous studies drew empirical evidence from Indo-European and Germanic languages such as English and Dutch. These languages are stress languages, which employ relative prominence between syllables (cued with salient pitch contours, lengthening, intensity increase, and vowel quality contrast; see Gordon & Roettger, 2017 for a review on cues of stress) to distinguish a limited number of word pairs (e.g., REcord and reCORD in English[1]). However, most of the world's languages are tonal languages, which use lexical tone (realized via pitch variation) to differentiate word meanings (Yip, 2002; Zsiga, 2012; Fromkin, 2014). For example, in Standard Chinese, a representative tonal language, the same segmental syllable *ma* means "mother" with a high-level tone (Tone 1, hereafter T1), "hemp" with a rising tone (Tone 2, hereafter T2), "horse" with a low dipping tone (Tone 3, hereafter T3), and "to scold" with a falling tone (Tone 4, hereafter T4). Therefore, to successfully recognize or produce a Standard Chinese word, it is crucial for Standard Chinese speakers to retrieve its corresponding lexical tone accurately and efficiently. Since most previous studies have focused on Western languages, the nature of lexical access in tonal languages, such as Mandarin, is not yet fully understood. For instance, the relative weighting and timing of utilizing segments versus lexical tone and the role of lexical tone in activating lexical candidates during the process of Mandarin spoken word recognition have remained controversial. Chapter 2 of this dissertation aimed to resolve the controversies with a series of eye-tracking visual world experiments.

The fact that many tonal language speakers are bilinguals further complicates the picture. As bilinguals activate words of both their languages

[1]Throughout the manuscript, we use capital letters to signal lexical stress.

during lexical access, one important issue that arises is the role of lexical tone in bilingual language co-activation. With bilinguals of two tonal systems, it is unclear whether their two tonal systems interact during lexical access, and if so, how they resolve the potential lexical conflicts. To investigate this issue, Chapter 3 of this dissertation studied the process of spoken word recognition with a unique type of bilinguals who speak two closely related dialects with mapping tones, namely Standard Chinese and Xi'an Mandarin bi-dialectals. As for bilinguals who speak both tonal and non-tonal languages, it is unclear whether lexical tone plays a role in non-tonal lexical access, especially during spoken word production. Are lexical tones activated even when bilinguals are speaking a non-tonal language? How does bilinguals' experience of speaking a tonal language affect non-tonal lexical access? With a series of picture-word interference tasks, Chapters 4 and 5 attempted to address these issues with Standard Chinese and English bilinguals.

In sum, to develop a more comprehensive account of lexical access, it is necessary to account for the role of lexical tone in both monolingual and bilingual speech comprehension and production. This dissertation aimed to fill in this gap by investigating the process of spoken word recognition and production in native speakers of Standard Chinese, bi-dialectal speakers of Standard Chinese and Xi'an Mandarin, and bilingual speakers of Standard Chinese and English. Specifically, we highlighted four issues: 1) the role of lexical tone in Mandarin spoken word recognition; 2) tonal interference in bi-dialectal spoken word recognition; 3) the activation of lexical tone in bilingual spoken word production; 4) the influence of lexical tone on bilingual mental lexicon. The rest of this chapter introduces each of the four issues and briefly explains how they are addressed using the visual world paradigm and the picture-word interference paradigm in this dissertation.

1.1 The Role of Lexical Tone in Mandarin Spoken Word Recognition

Several studies have shown activation and competition of phonologically similar words during the process of spoken word recognition in Mandarin (e.g., Lee, 2007; Liu & Samuel, 2007; Malins & Joanisse, 2010; Sereno & Lee, 2010; Zhao et al., 2011), similar to Indo-European languages. However, a few issues concerning how exactly segmental and tonal cues are taken up and processed remain open. First, it is controversial whether segmental syllables have a special status in Mandarin lexical processing. Different from syllables of Indo-European languages, Mandarin syllables are simpler in structure and limited in types and numbers (e.g., Roelofs, 2015; Verdonschot et al., 2015; Chen & Chen, 2002); as most Mandarin morphemes are monosyllabic and the writing system is based on syllable-sized characters, researchers have entertained the idea that the (segmental) syllable is a holistic processing unit in Mandarin lexical access (e.g., Zhao et al., 2011; Sereno & Lee, 2016). However, experimental data on this issue are rather controversial (e.g., see Zhao et al., 2011 and Sereno & Lee, 2016 for evidence supporting a special status of the segmental syllable; but see Malins & Joanisse, 2012 for evidence against it). Second, existing studies on Mandarin spoken word recognition differ on whether and to what extent sub-lexical components such as onset, rhyme and lexical tone affect lexical activation. In Indo-European languages, word candidates with the same onset are generally activated earlier and greater than word candidates with the same rhyme. For instance, with the visual world paradigm, Allopenna, Magnuson, and Tanenhaus (1998) found that when listening to targets (e.g., “beaker”), while both cohort competitors (e.g., “beetle”) and rhyme competitors (e.g., “speaker”) drew more of listeners’ attention than unrelated distractors (e.g., “carriage”), listeners’ eye fixations towards cohort competitors were significantly earlier than those of rhyme competitors. However, the effects of cohort and rhyme competitors are found to be less reliable in Mandarin spoken word recognition. For instance, using the visual world paradigm, Malins and Joanisse (2010) found a null effect of rhyme competitors whereas Zou

(2017) found a null effect of cohort competitors. Third, the time course of utilizing segmental and tonal cues during Mandarin spoken word recognition is not clear. Previous studies have reported a perceptual disadvantage of lexical tone compared with segmental information (Taft & Chen, 1992; Yip, 2001; Cutler & Chen, 1997; Ye & Connine, 1999, experiment 1; Hu et al. 2012; Sereno & Lee 2015; Gao et al., 2019). Such a disadvantage is often attributed to the fact that tonal information that arrives later and thus is processed later than segmental information (Cutler & Chen, 1997). However, recent evidence from eye-tracking and event-related potentials (ERPs) data show that tonal information plays an equivalent role to segmental information, and it is processed timely during Mandarin spoken word recognition (Malins & Joanisse, 2010; 2012). Questions about the time course of segmental and tonal processing and cue utilization during Mandarin lexical processing thus remain to be answered.

We addressed these issues in Chapter 2 with three eye-tracking visual world paradigm experiments. We aimed to clarify the role of segmental syllable and sub-syllabic constituents, as well as to investigate the time course of using segmental and suprasegmental tonal information during Mandarin lexical processing. In Experiments 1 and 2, native Standard Chinese (hereafter SC) speakers listened to monosyllabic SC words with the presence of a phonological competitor, which overlaps with the target in either segmental syllable, onset and tone, rhyme and tone, or just tone. Experiments 1 and 2 differ in how long listeners were allowed to preview pictures on the screen before hearing the spoken target word, as previous studies found that the length of preview time plays a crucial role in observing phonological competition effects or not (e.g., Huettig & McQueen, 2007; Huettig et al., 2011). Eye movement results of both Experiments 1 and 2 confirmed a robust competition effect of segmental syllable overlap competitors, and null effects of onset, rhyme and tone overlap distractors. Experiment 3 investigated the time course of segmental versus tonal information utilization by manipulating their point of divergence in acoustic cues. We found that both sub-syllabic information (i.e., segment vs. tone) and cue timing (i.e.,

early vs. late point of divergence) affect phonological competition effects. Regardless of the nature of the cues, the point of divergence determines the amplitude and time course of the competition effect: the earlier the point of divergence, the sooner the competition, suggesting that despite the dominant role of the segmental syllable, Mandarin listeners use both segmental and tonal information as soon as they are available to constrain lexical activation.

1.2 Tonal Interference in Bi-dialectal Spoken Word Recognition

In Chapter 3, we investigated the process of spoken word recognition in bi-dialectal speakers of Standard Chinese and Xi'an Mandarin. Both Standard Chinese and Xi'an Mandarin belong to the Mandarin Chinese family. They share similar syntactic structures, a large number of etymologically related translation equivalents, the same writing system and nearly the same segmental inventories. Moreover, the lexical tone systems of Standard Chinese and Xi'an Mandarin have a one-to-one mapping relation (Liu et al., 2020), resulting in a large number of cross-dialect homophones across the two dialects. For example, *ma* with a high-level tone means "mother" in SC, whereas it means "to scold*"* in Xi'an Mandarin. Such a unique case of Standard Chinese and Xi'an Mandarin bi-dialectals provides us with the opportunity to test whether and to what extent the mapping of tonal systems elicits cross-dialect interference while keeping the orthographic, morphological, semantic, and segmental aspects constant.

Using generalized lexical decision tasks with auditory priming, Liu (2018) manipulated five types of target and prime contrasts based on the cross-dialect phonological similarity between Standard Chinese and Xi'an Mandarin: 1) within-dialect segment and tone overlapping target and prime; 2) within-dialect segment overlapping target and prime; 3) cross-dialect segment and tone overlapping target and prime (i.e., cross-dialect homophones); 4) cross-dialect segment overlapping target and prime; 5) unrelated target and prime. Results of reaction times showed that with Standard Chinese primes, there was a significant

interference effect for the cross-dialect segment and tone overlapping targets but not for the cross-dialect segment overlapping targets, compared with unrelated targets. Liu (2018) interpreted these results as evidence for cross-dialect tonal interference during bi-dialectal spoken word recognition. Note that this effect was found in a mixed-dialectal context, i.e., bi-dialectals were exposed to words in both Standard Chinese and Xi'an Mandarin, which might have created or boosted cross-dialectal interference. Using the eye-tracking visual world paradigm, Chapter 3 sought to further understand the effects of tonal interference in a controlled mono-dialectal context. Specifically, we asked Standard Chinese and Xi'an Mandarin bi-dialectals to listen to sentences in one dialect and identify the target word among four Chinese characters shown on the screen. The characters included the target, two unrelated distractors, and a phonological competitor which share the same segmental syllable with the target within and across dialects. Among the phonological competitors, besides segmentally overlapping distractors which does not share lexical tone with the target within and across dialects (Segment Condition), there were also cross-dialect homophone competitors that share the same lexical tone with the target across dialects (Homophone Condition) and translation-induced cross-dialect homophones that share the same lexical tone with the targets' dialectal translation equivalent (Translation Condition). We hypothesized that, if both sets of lexical tones are activated, the Homophone and Translation Condition would elicit larger competition effects than Segment Condition; if only one set of lexical tones are activated, Segment Condition would elicit most competition effects, because the tonal contours of the target and competitor of the Segment Condition share most acoustic similarity. Listeners' eye movements show that distractors in the Segment Condition interfere with participants' eye fixations significantly more than Homophone and Translation Conditions, suggesting a lack of cross-dialectal interference effect. It is likely that the mono-dialectal sentence context has cancelled out the cross-dialect interference effect shown in Liu (2018). Overall, this finding marks a convergence between bi-dialectal and bilingual speech

processing. Based on these findings, a preliminary model of bi-dialectal spoken word recognition which emphasises active control of dialect activation was proposed.

1.3 The Role of Lexical Tone in Bilingual Spoken Word Production

Bilinguals not only retrieve the form of the target language but also that of the non-target language during spoken word production (See Costa, 2009 for a review). For example, with the picture-word interference paradigm, Dutch-English bilinguals were found to take longer to name pictures in their L2 English when the Dutch auditory distractor was phonologically similar to the Dutch translation of the target picture, compared with unrelated distractors (e.g., target *berg* "mountain" – phono-translation distractor *berm* "verge" – unrelated distractor *kaars* "candle"; Hermans et al., 1998). This finding indicates that the translations of the non-target language are activated at the phonological level. However, most previous studies on bilingual word production have focused on segments. It remains open whether suprasegmental information such as lexical tone is co-activated during bilingual spoken word production.

In Chapter 4, we aimed to address this issue by examining the role of lexical tone in English spoken word production with bilinguals of Standard Chinese and English. Specifically, we asked: if Standard Chinese and English bilinguals co-activate both Standard Chinese and English names during English word production, is lexical tone co-activated and utilized during the process? With four picture-word interference experiments, Standard Chinese and English bilingual speakers were instructed to name pictures in English (e.g., feather) while ignoring four types of simultaneously presented SC distractors: 1) the translation distractor, which is the translation equivalent of the English target name (e.g., *yu3mao2* "feather"); 2) the tone-sharing distractor, which shares both tone and segments with the SC translation in the first syllable (e.g., *yu3zhou4* "universe"); 3) the no-tone-sharing distractor, which shares segments only with the Standard

Chinese translation in the first syllable (e.g., *yu4mi3* "corn"); 4) the unrelated distractor, which shares no phonological overlap with target and its translation (e.g., *lei4shui3* "tear"). To further explore potential factors that may constrain the lexical tone effect, we also manipulated two additional factors that have been found to affect picture naming onset with the picture-word interference paradigm. One was distractor modality (e.g., Hantsch et al., 2009; Jonen et al., 2021); the SC distractors were presented either auditorily or visually. The other was familiarization mode (e.g., Llorens et al., 2014); bilinguals were asked to familiarize with the target pictures' English names only (i.e., English mode) or both English and Standard Chinese names (i.e., mixed mode). In Experiment 1 (with auditory distractor and English mode), translation distractors significantly facilitated bilingual English picture naming, while tone-sharing distractors significantly inhibited the process. Importantly, the tone-sharing distractors elicited significantly longer naming latency than the no-tone-sharing distractors, demonstrating the co-activation of lexical tone during English spoken word production. Overall, this study replicated previously found translation facilitation effect (e.g., Costa et al., 1999) and observed a significant interference effect of lexical tone. These findings suggest that Standard Chinese and English bilinguals not only co-activate the Standard Chinese translation equivalents but also the lexical tones of the Standard Chinese translations during English spoken word production. Results of Experiment 2 (auditory distractor and mixed mode), Experiment 3 (visual distractor and English mode), and Experiment 4 (visual distractor and mixed mode) further demonstrated that the polarity and robustness of the lexical tone effect are modulated by external factors such as distractor modality and familiarization mode.

1.4 The Effect of Lexical Tone on the Bilingual Mental Lexicon

Although it is widely agreed that words of bilinguals' two languages interact in their mental lexicon (see Kroll et al., 2012 for a review), how suprasegmental

features interact is still unclear, especially for bilinguals of two typologically different languages such as Standard Chinese and English. In Chapter 5, we further asked whether and to what extent lexical tone modulates pitch processing in non-tonal speech production with Standard Chinese and English bilinguals.

Previous studies on bilingual lexical access have identified an important distinction between Standard Chinese and English bilinguals and native English speakers in pitch processing during spoken word recognition (Ortega-Llebaria et al., 2017; 2020). With primed-lexical decision tasks, Ortega-Llebaria et al. (2020) asked Standard Chinese and English bilinguals and English monolinguals to make lexical judgments on English target words produced in either falling or rising pitch contours, while the prime and target were manipulated to fully match, fully mismatch, mismatch in segments, or mismatch in pitch. Results showed that in the full match and pitch mismatch conditions, Standard Chinese and English bilinguals experienced larger facilitation when the targets were produced with a falling pitch contour than with a rising pitch contour. Yet, such a "falling-f0 bias" was only found in SC-English bilinguals but not English monolinguals. This fact has led Ortega-Llebaria et al. (2017; 2020) to reason that words with falling pitch contours are closer English lexical representations than their rising counterparts in the mental lexicon of SC-English bilinguals; crucially, it is the long-term experience with lexical tone that reshapes their mental lexicon. This assumption is also consistent with the observation that Standard Chinese learners of English tend to produce stressed syllables with an H* pitch accent, giving the English words a falling-like pitch contour (McGory, 1997). However, no study so far has directly tested the effect of lexical tone on pitch representation and processing in bilingual spoken word production.

In Chapter 5, we adopted the picture-word interference paradigm to investigate this issue. Previous bilingual picture-naming studies have shown that cross-language homophone distractors facilitate picture naming in non-target languages (e.g., Hermans et al., 1998; Costa et al., 1999, 2003). In this study, we asked Standard Chinese and English bilinguals and native English monolinguals

to name pictures in English (e.g., *lung*) while ignoring simultaneously played SC cross-language homophones that either have a falling or a rising lexical tone (*lang4* with a falling tone, "wave"; *lang2* with a rising tone, "wolf"). We hypothesized that if lexical tone indeed influences bilinguals' pitch representation in non-tonal second languages, the effect of lexical tone (falling vs. rising) on English picture naming should differ between Standard Chinese and English bilingual and English monolingual speakers. Results showed that, compared with unrelated Standard Chinese distractors, both falling and rising cross-language homophones facilitated English word naming for both SC-English bilingual and English monolingual speakers. Most importantly, SC-English bilinguals showed significantly longer naming latencies with falling-tone in cross-language homophones than their rising-tone counterparts, whereas English monolingual speakers did not show such a pattern. As one of the first studies that investigated the influence of lexical tone in non-tonal lexical access during spoken word production, we identified a significant difference between SC-English bilinguals and English monolinguals in terms of how falling versus rising lexical tones affect English picture-word naming. This finding provides important implications for understanding pitch representation and processing in the bilingual lexicon, as well as the interaction between bilinguals' two languages at the suprasegmental level.