Lexical tone in word activation

Yang, Q.

Chapter 6

General Discussion

This dissertation investigates the process of spoken word recognition and spoken word production in native speakers of Standard Chinese, bi-dialectals of Standard Chinese and Xi'an Mandarin, and bilinguals of Standard Chinese and English. While most previous studies on lexical processing focused on the use of segmental information, this dissertation provides important complementary evidence with data on suprasegmental information, i.e., lexical tone, which can help us develop a more comprehensive account of (bilingual) lexical access. In the following sections, the findings of each chapter, the general implications of the findings and future directions are summarized.

6.1 Chapter-by-chapter Summary

Chapter 2 aimed to resolve three controversial issues in Mandarin spoken word recognition: 1) Do segmental syllables have a special status in Mandarin lexical processing? 2) What are the relative contributions of onset, rhyme, and lexical tone? 3) What is the time course of segmental and tonal processing during online lexical processing? To address these questions, three eye-tracking visual world paradigm experiments were conducted. Experiments 1 and 2 examined the relative contribution of the segmental syllable, onset, rhyme, and lexical tone in Mandarin lexical processing by investigating to what extent participants' visual attention is distracted by the presence of competitors during the process of recognizing the target spoken word. Critically, five types of competitors were manipulated based on their phonological overlap with the target, namely, segmental syllable competitors (with segmental syllable overlap), cohort competitors (with the onset and lexical tone overlap), rhyme competitors (with rhyme and lexical tone overlap), tonal competitors (with lexical tone overlap), and unrelated distractors (with no overlap). While Experiment 1 allowed participants to preview the pictures for 1,500 ms before listening to the target word, Experiment 2 had a shorter preview of 200 ms. Both experiments found that only segmental syllable competitors significantly distracted participants' visual

attention towards the target word more than unrelated distractors. Cohort competitors, rhyme competitors, and tonal competitors did not affect participants' visual attention more than unrelated distractors. Experiment 3 zoomed further into listeners' sensitivity to the acoustic details of segmental and tonal information. The target and competitor differed in either segmental or tonal information. Moreover, we manipulated the point of information divergence (early vs. late) between the target and competitor word pair along both segmental and tonal dimensions. Specifically, while the tonal early diverging target and competitor only share the same segments, the tonal late diverging pair share the same segment and the onset of the tonal pitch contours; while the segmental early diverging target and competitors share the onset and lexical tone, the segmental early diverging pair share the onset, glide, and lexical tone. Eye-tracking results show that, while both tonal early and late diverging competitors significantly attracted participants' visual attention, the late competitors exhibited significantly larger effects than the early competitors. Moreover, no statistically significant difference was found between tonal and segmental competitors, regardless of the point of divergence. In sum, the results of Experiments 1 and 2 indicate an advantageous role of segmental syllable over onset, rhyme, and lexical tone in activating word candidates; Experiment 3 shows that both tonal and segmental information can be used as soon as they are available to constrain word candidates' activation during the process of Mandarin spoken word recognition.

In Chapter 3, we further questioned whether and to what extent two tonal systems interact in listeners of two closely related tonal dialects. To answer these questions, we investigated the process of spoken word recognition with bi-dialectal speakers of Standard Chinese and Xi'an Mandarin. With the visual world paradigm, Standard Chinese and Xi'an Mandarin bi-dialectal speakers were asked to listen to short sentences produced in either Standard Chinese or Xi'an Mandarin (e.g., *wo3 yao4 shuo1 hua1*; "I will say flower") and identify the target word (e.g., *hua1* "flower") among four Chinese characters shown on the computer screen. The four characters included the target, two unrelated distractors, and a

phonological competitor. All phonological competitors share the same segmental syllable with the target within- and cross-dialects. Among the phonological competitors, there were cross-dialect homophone competitors that share the same lexical tone with the target across dialects (Homophone Condition), translation-induced cross-dialect homophones that share the same lexical tone with the targets' dialectal translation equivalent (Translation Condition), and competitor that does not share lexical tone with the target either within- or cross-dialects (Segment Condition). We hypothesized that, if both sets of lexical tones are activated, (translation-induced) cross-dialect homophones would elicit larger competition effects than competitors that have segmental overlap with targets only (the Segment Condition). Results of Standard Chinese and Xi'an Mandarin bi-dialectals' eye movements showed that, regardless of listening in either Standard Chinese or Xi'an Mandarin, neither the Homophone nor the Translation Condition distracted participants' eye fixations more than the Segment Condition competitors. Rather, the Segment Condition exhibited a larger phonological competition effect than the Homophone and Translation Conditions due to larger tonal similarities between the target and competitor word pairs within one dialect. Overall, this finding suggests a lack of lexical co-activation across dialects of Xi'an Mandarin and Standard Chinese. Given that previous studies on bilingualism have found consistent evidence that bilinguals co-activate both their languages during spoken word recognition with similar experimental set-ups (e.g., Spivey & Marian, 1999; Shook & Marian, 2017; Wang, Wang & Malins, 2017), our results indicate a lexical processing divergence between bilingual and bi-dialectal speech comprehension. Based on these findings, we proposed a preliminary bi-dialectal spoken word recognition model that emphasises the dialect control mechanism.

In Chapter 4, we continued to explore the role of lexical tone in bilingual spoken word production. Specifically, we asked whether Standard Chinese and English bilingual speakers co-activate lexical tone even when producing an English spoken word. With picture-word interference experiments (Rosinski et al.,

1975), Standard Chinese and English bilingual speakers were asked to name pictures in English (e.g., feather) while ignoring four types of simultaneously presented Standard Chinese distractors: 1) the translation distractor, which was 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 SC 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"). Moreover, we manipulated the distractor modality and the familiarization mode before the naming task. Specifically, Standard Chinese distractors were presented auditorily in Experiments 1 and 2, but visually in Experiments 3 and 4. Before performing the naming task, bilinguals were familiarized with the target pictures' English names in Experiments 1 and 3, whereas both English and Standard Chinese names in Experiments 2 and 4. Results in Experiment 1 (auditory distractor and English mode) showed that translation distractors significantly facilitated target picture naming but tone-sharing distractors significantly interfered with target picture naming. Moreover, there was a significant naming latency difference between the tone-sharing and no-tone-sharing distractors, demonstrating the co-activation of lexical tone during English spoken word production. In Experiment 2 (auditory distractor and mixed mode), translation and phono-translation distractors all elicited interference towards target naming, but none of the effects was statistically significant. In Experiment 3 (visual distractor and English mode), there was a robust translation facilitation effect; the tone-sharing and no-tone-sharing distractors were also found to be facilitatory, but neither effect was statistically significant. In Experiment 4 (visual distractor and mixed mode), there was also a robust translation facilitation effect; the tone-sharing distractors significantly facilitated picture naming whereas the no-tone-sharing distractors did not, indicating an important role of lexical tone during English picture naming. Overall, replicating previously identified translation facilitation effects (e.g.,

Costa et al., 1999), this study discovers a significant difference between the tone-sharing and no-tone-sharing conditions. 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. Moreover, the polarity and robustness of the lexical tone effect in spoken word production are modulated by procedural factors such as the distractor modality and the familiarization mode.

In Chapter 5, we investigated the influence of lexical tone on pitch representation and processing during non-tonal word production with Standard Chinese and English bilingual speakers. With the picture-word interference paradigm, we asked Standard Chinese and English bilinguals and native English monolinguals to name pictures in English (e.g., *lung*) while ignoring simultaneously played Standard Chinese cross-language homophones that either have a falling or a rising lexical tone (*lang* with a falling tone, “wave”; *lang* with a rising tone, “wolf”). We hypothesized that if lexical tone indeed affects bilinguals’ pitch representation in L2 non-tonal languages, the effect of lexical tone (falling vs. rising) on English picture naming should differ between Standard Chinese and English bilinguals and English monolingual speakers. Naming onset results show that, while both falling and rising cross-language homophones facilitated English word naming, only bilinguals of Standard Chinese and English, but not English monolingual speakers, showed significantly longer naming latencies with falling-tone cross-language homophones than their rising-tone counterparts. Such a distinction between Standard Chinese and English bilinguals and English monolinguals suggests that lexical tone plays an important role in pitch representation and processing during bilingual non-tonal spoken word production.

6.2 Theoretical Implications

Drawing empirical evidence from native Standard Chinese speakers, bi-dialectal speakers of Standard Chinese and Xi'an Mandarin, and bilingual speakers of Standard Chinese and English, our findings on spoken word recognition and production highlight the role of lexical tone during word co-activation and competition within- and across-languages. Current models of (bilingual) spoken word recognition and production should be adjusted to account for the possibilities of tonal processing. Moreover, findings on "tonal bilinguals" (Wu, 2015) of two closely related dialects should also be taken into consideration in our understanding of how the two linguistic systems of a speaker may interact in dynamically different ways.

Although previous studies have made a few attempts to modify current models to account for tonal word recognition (e.g., Ye & Connie, 1999; Yue, 2016; Gao et al., 2019; Shuai & Malins, 2017; Tong et al., 2014; Zhao et al., 2011), they disagree on whether an extra level of a syllable or segmental syllable is necessary (e.g., Zhao et al., 2011; Yue, 2016; Gao et al., 2019), and whether segment and tone processing are integrated (e.g., the TTRACE model; Tong et al., 2014) or separated (e.g., the TRACE-T model; Shuai & Malins, 2017). Given that in Chapter 2, we found an advantageous role of segmental syllables over sub-lexical constitutes such as onset and rhyme, and that both tonal and segmental information were used incrementally in Mandarin spoken word recognition, we proposed a revised TRACE model for tonal word recognition with a four-layer structure: syllable, segmental syllable, phonemes, and lexical tone. The extra level of segmental syllable accounts for the overall larger and more stable phonological competition effects of segmental syllable over a combination of sub-lexical phonological components during Mandarin spoken word recognition. Moreover, with independent representations of phonemes and tones, both phonemic and tonal information can be used to resolve phonological competition as soon as they are available.

While findings in Chapter 2 provide important implications for models of spoken word recognition in a lexical tone language, our findings on bilingual spoken word production in Chapters 4 and 5 further shed light on our understanding of the role of lexical tone in spoken word production. In general, current models of spoken word production either support the early active selection of lexical tone (e.g., Wan & Jaeger, 1998; Alderete et al., 2019) or not (e.g., Roelofs, 2015). According to the former view, the lexical tone is represented independently at the same operational level of segments and can be actively selected at an early stage of phonological encoding (Alderete et al., 2019). The latter view adapts the influential WEAVER++ model (Roelofs, 2000) to Mandarin in positing that lexical tone is represented diacritically as a metric frame, and only associated with pre-selected segments at a late stage of phonetic spell-out. In Chapter 4, we found that even during non-tonal English word production, lexical tone is co-activated and plays an important role in cross-language competition and selection, which is unlikely the case if lexical tone is only implemented during phonetic spell-out. In Chapter 5, we found further evidence that lexical tone directly modulates pitch processing in English spoken production, which in itself argues for a more influential role of lexical tone than mere labels in the metric frame. Overall, our findings strengthen the independent view of lexical tone selection in spoken word production from a bilingual lexical access perspective. Simultaneously, they validate the discrete and interactive processing of sub-lexical and lexical representations in speech production from a broader perspective on lexical access.

As previous studies either focused on bilingual or monolingual language processing, the nature of bi-dialectalism has been left largely unresearched. It has been debated whether bi-dialectal language processing should resemble that of bilinguals or monolinguals (see Melinger, 2018 for a review of the two views of bi-dialectalism). Chapter 3 focused on speakers of two closely related tonal dialects, Standard Chinese and Xi'an Mandarin, which provide important implications for our understanding of bi-dialectalism and lexical tone interaction.

Based on the finding that, unlike bilinguals, bi-dialectals of Standard Chinese and Xi'an Mandarin were able to achieve selective access to the target dialect, we proposed a spoken word recognition model for bi-dialectals. This model includes levels of phonological, phono-lexical, ortho-lexical, and semantic representations; within each level, dialect-specific and dialect-shared features are stored in the same space, allowing communication and competition between dialects. Moreover, there is a task scheme in the model that functions as a dialectal membership tag. By making use of the environment and task requirements, the task scheme can suppress the entire lexicon of the non-target dialect. This basic framework of a preliminarily verbal model is inspired by previous bilingual comprehension models such as BLINCS (Shook & Marian, 2013), BIA (e.g., Grainger & Dijkstra, 1992; Dijkstra et al., 1998), BIA+ (Dijkstra & Heuven, 2002) and further extends to account for bi-dialectal lexical processing.

6.3 Methodological Contributions

This dissertation also has a few methodological contributions to the implementation of the visual world paradigm and the picture-word interference paradigm.

Two factors in the visual world paradigm have been long questioned regarding the extent to which they affect the eye-tracking results (see reviews in Huettig et al., 2011; Apfelbaum et al., 2021). One is the length of the preview time, i.e., the time allowed for participants to view the pictures on screen before listening to the auditory stimuli. It has been proposed that a short preview time such as 200 ms is not sufficient for participants to retrieve the names of the displayed pictures and thus results in null phonological competition (e.g., Huettig & McQueen, 2007); however, there have also been studies that found effects of phonological competition even without any preview time (e.g., Zou, 2017). In Chapter 2, with a short preview time of 200 ms, we replicated results obtained with a long preview time of 1,500 ms in finding a robust segmental syllable

phonological competition effect. Besides, we found that listeners located the target picture faster with fewer eye fixations with the short preview time than with the long one, along with subtle differences in the time course of eye fixations. These findings suggest that although the length of preview time is not a determining factor for observing phonological competition, it does affect how listeners distribute their visual attention. The other factor is the use of Chinese characters as visual displays. While previous studies have validated using printed words to study spoken word recognition in Western languages (Huettig et al., 2011), it is less clear to what extent the display of Chinese characters affects Mandarin lexical processing in the visual world paradigm. In Chapter 2, we found a subtle trend of cohort competition with the display of Chinese characters, which was missing from the picture display. This discovery indicates that the incorporation of Chinese characters into the picture world paradigm not only confirms its validity but also demonstrates an increased sensitivity to phonological effects.

With the picture word interference paradigm, we also manipulated two procedural factors. One is the modality of the distractors, i.e., whether participants listened to or viewed distractor words during picture naming (auditory vs. visual distractors). The other is the familiarization mode of the target pictures, i.e., whether participants were given English names only (i.e., the English mode) or both English and SC names (i.e., the mixed mode) during the familiarization session. While both factors have been found to influence the naming process in the picture world interference paradigm (e.g., Hantsch et al., 2009; Llorens et al., 2014; Jonen et al., 2021), it has not been answered whether and if so, to what extent the two factors affect cross-language co-activation and competition in bilingual studies. For instance, it is unknown whether distractor modality might be responsible for causing opposite effects of translation and phono-translation distractors, given that previous studies mostly observed translation facilitation with visual distractors (e.g., Costa et al., 1999; Costa & Caramazza, 1999; Hermans, 2004) and phono-translation interference with auditory distractors (e.g.,

Hermans et al., 1998; Costa et al., 2003). Our findings in Chapter 4 found that, while both auditory and visual translation distractors could facilitate bilingual picture naming, only auditory phono-translation distractors, but not their visual counterparts, significantly interfered with the process. The familiarization mode was also found to have an impact on the effect of language co-activation: familiarizing target pictures' names in both languages significantly reduced the translation facilitation effect of auditory distractors, compared with familiarizing names in the target language alone. Future spoken word production studies, therefore, should take both factors into account when interpreting the direction and robustness of the cross-language effects within the (bilingual) picture-word interference paradigm. Moreover, Chapters 4 and 5 were conducted online due to the influence of COVID-19. By successfully replicating previous lab findings such as the translation facilitation effect (e.g., Costa et al., 1999), we showed that implementing the picture-word interference paradigm online is an efficient and sound approach to studying the process of spoken word production.

6.4 Limitations and Future Directions

First, further research is needed to investigate the relationship between tonal word recognition and production in Mandarin. Our findings in Chapter 2 indicate that segmental syllables play a primary role in spoken word recognition, which is consistent with previous studies on Mandarin word production (Meyer, 1991; Chen, Lin, & Ferrand, 2003; Chen & Chen, 2013; Chen, O'Seaghdha & Chen, 2016; Wang, Wong, & Chen, 2018). In our proposed revised TRACE model, we incorporated an extra level of segmental syllable and independent representation of lexical tones, which aligns with the idea in speech production that the atonal syllable is "the proximate phonological encoding" (O'Seaghdha, 2010; Roelofs, 2015). However, potential asymmetries and differential engagement of segmental syllables and lexical tone in tonal word perception and

production remain unexplored. Therefore, further studies are needed to address these questions.

Second, in Chapter 3, we proposed a model of bi-dialectal lexical access that emphasizes potential differences in the control mechanism between bi-dialectals and bilinguals. We also hypothesized that the differences between bi-dialectal and bilingual lexical access may vary along a continuum, depending on the degree of similarities between the dialects. However, further studies on other dialects and pairs of languages with typologically different prosodic systems are still needed to validate our models and to explore the extent to which bilingual and bi-dialectal lexical processing differ.

Third, although we explained our findings in spoken word recognition and production within the framework of interactive activation and competition (e.g., McClelland & Rumelhart, 1981; Chen & Mirman, 2012), it is worth noting that other theories, such as the response exclusion hypothesis (Mahon et al., 2007) may also account for our findings without referring to lexical competition. In a similar vein, although we mainly explained our bilingual findings within the framework of language co-activation, it is important to note that the learning account (e.g., Costa et al., 2017), which does not resort to language co-activation, may also explain some of our findings. Although the question of whether lexical selection depends on co-activation and competition is a compelling research topic, we did not delve into it further in this dissertation for lack of judicating evidence in our results.

Last but not least, we proposed various modifications of current models and theories of lexical processing to account for tonal word recognition and production, as well as bi-dialectal lexical access. However, our models and suggestions have remained verbal and preliminary. To validate our models, computational simulations and additional empirical evidence from various tonal languages and dialects are necessary.