Lexical tone in word activation
Yang, Q.

Chapter 4

The Role of Lexical Tone in Bilingual Spoken Word Production

Abstract

During spoken word production, bilinguals not only retrieve the form of the target language but also that of the non-target language. However, most previous studies on bilingual word production have focused on segments. It remains open whether suprasegmental information is involved in the process as well. We aimed to address this gap by examining the role of lexical tone in English spoken word production with bilinguals of Standard Chinese (hereafter SC) and English. In four online picture-word interference (PWI) experiments, we asked SC-English bilingual speakers to name pictures in English (e.g., feather) while ignoring four types of simultaneously presented SC distractors: 1) a translation distractor, i.e., the translation equivalent of the English target name (e.g., *yu3mao2* "feather"); 2) a tone-sharing distractor, which shares both tone and segments with the SC translation in the first syllable (e.g., *yu3zhou4* "universe"); 3) a no-tone-sharing distractor, which shares segments but not tone with the SC translation in the first syllable (e.g., *yu4mi3* "corn"); 4) an unrelated distractor, which shares no phonological overlap with the target or its translation (e.g., *lei4shui3* "tear"). We also manipulated two procedural factors, namely distractor modality (i.e., whether distractors were presented auditorily or visually) and familiarization mode (i.e., whether participants previewed the target picture names in English or in both English and SC). Not only did this study replicate the translation facilitation effect (e.g., Costa et al., 1999) but also observed significant differences between the effects of tone-sharing and no-tone-sharing distractors. Moreover, the polarity and robustness of such effects are subject to the interaction of distractor modality and familiarization mode. Overall, our findings suggest that SC-English bilinguals co-activate the lexical tone of SC translations during English picture naming.

Keywords: lexical tone; spoken word production; language co-activation

One of the most important findings in the bilingual literature is that bilinguals co-activate both of their languages even when they are only using one. There is substantial evidence of co-activation not only for spoken word comprehension (e.g., Dijkstra & Van Heuven, 2002; Thierry & Wu, 2007; Kroll & De Groot, 2009) but also for spoken word production (see Costa, 2009 for a review). It is important to note that while the literature on spoken comprehension shows that not only segments but also suprasegmental features such as lexical tone are co-activated (e.g., Wang et al., 2017), studies on bilingual spoken word production have mainly focused on segments (e.g., Colomé, 2001; Costa & Caramazza, 1999; Hermans et al., 1998). The goal of this study is to fill in this knowledge gap and investigate whether suprasegmental information of the non-target language is also co-activated during the process of spoken word production. More specifically, we will address this question by examining the co-activation of lexical tone in English and Mandarin bilingual spoken word production within the picture-word interference paradigm (hereafter PWI; Rosinski et al., 1975).

PWI is one of the most widely used paradigms for examining the process of spoken word production. In this paradigm, participants are asked to name a picture while ignoring the presence of a written or spoken distractor word. It generally takes participants more time to name the target picture if the name of the target picture and the distractor word are semantically related (e.g., target *dog* – distractor *cat*), and less time if they are phonologically related (e.g., target *dog* – distractor *doll*), compared with an unrelated condition. These effects are known as the *semantic interference* effect and *phonological facilitation* effect, respectively. The rationale behind the effects is that the retrieval of the target picture's concept not only activates the target word (e.g., *dog*) but also words that are semantically related to the target (e.g., *cat*). Thus, a semantic distractor (e.g., *cat*) would receive activation from both the target picture and the distractor word. Compared with unrelated distractors (e.g., *table*), which only receive activation from the distractor word itself, the activation level of the semantic distractor is thus higher and interferes more with target selection (but see Mahon et al., 2007 for a different account). A phonological distractor (e.g., *doll*), on the other hand, facilitates picture naming because their shared phonological properties aid the retrieval of the targets' phonological form.

Interestingly, in a renowned bilingual study by Costa et al. (1999), *semantic interference* and *phonological facilitation* effects were also found across languages. When Spanish and Catalan bilinguals were asked to name pictures in Spanish, Catalan distractors elicited *semantic interference* and *phonological facilitation* just like Spanish distractors, suggesting that lexical activation is not limited to the target language. In that study, Costa and his colleagues also found that the Catalan translations of the targets significantly facilitated Spanish picture naming. This effect, commonly known as *translation facilitation* or between-language identity effect, has been taken as a strong indicator of bilingual language co-activation (e.g., Costa & Caramazza, 1999; Hermans, 2004). Moreover, a translation-mediated phonological effect has been found in bilingual PWI studies. For instance, in Hermans et al. (1998), Dutch-English bilinguals took longer to name pictures in their L2 English when the auditory Dutch distractor (e.g., *berm* "verge") is phonologically similar to the Dutch translation of the picture (e.g., *berg* "mountain") compared with unrelated distractors (e.g., *kaars* "candle"). Such an effect, known as the *phono-translation interference* effect, has been replicated with bilinguals of high proficiency in both their languages (Costa et al., 2003), bilinguals of typologically distant languages (French and Tunisian Arabic; Boukadi et al., 2015), and in bilinguals' native language (Klaus et al., 2018). It indicates that not only the lexical representations, but also the sub-lexical phonological representations of the translation equivalents are co-activated during speech production. The *phono-translation interference* effect also serves as evidence supporting cascaded theories of lexical access, which argues that phonological and lexical representations are linked through cascaded and interactive processing (see Schiller & Alario, 2023 for a review on cascaded models of speech production).

In addition to the *phono-translation interference* effect in PWI, evidence indicates that bilinguals co-activate the phonological representations of both languages in various speech planning and production tasks. For example, with a phoneme monitoring task, Colomé (2001) found that when Catalan-Spanish bilinguals were asked to detect certain phonemes in the Catalan name of a picture (e.g., *taula*, “table”), they took longer to reject phonemes (e.g., /m/) that are in the target pictures’ Spanish translation (e.g., *mesa*) than those that are not (e.g., /f/). Similarly, Macizo (2015) asked Spanish-English bilinguals to name the colour of a pictured object in English (e.g., *brown suitcase*) and found that they took longer to respond when the names of the colour and picture were phonologically related in Spanish (e.g., *maleta marrón*, “brown suitcase”) compared with unrelated counterparts (e.g., *maleta rosa*, “pink suitcase”). Consistently, evidence from bi-dialectal processing supports the same view: using an auditory lexical decision experiment, Wu et al., (2015) observed that cross-dialect tonal similarity between SC and Jinan Mandarin manipulates SC-Jinan Mandarin bi-dialectals’ lexical processing of etymologically-related translation equivalents.

Despite the substantial evidence that bilinguals have access to the phonology of both languages during spoken word production (e.g., Hermans et al., 1998; Costa et al., 2000; Colomé, 2001; Roelofs, 2003; Macizo, 2016; Spalek et al., 2014; Klaus et al., 2018a), it is important to note that the existing studies mainly drew evidence from segmental information. Few studies have looked into the co-activation of suprasegmental information in bilingual word production. The only study to our knowledge is Martínez García (2018), which compared the co-activation of stress-sharing and no-stress-sharing cognates in English-Spanish bilinguals. In this study, participants were asked to name a printed Spanish word (e.g., *materia* “subject”) while ignoring an English cognate competitor that was also in the display (e.g., *material* “material”). Critically, the competitor either shares the same stress pattern with the target (e.g., target *maTEria* “subject” – competitor *maTErial*) or not (e.g., target *liTEra* “bunk bed” – competitor *LIteral*). By comparing the naming latencies, Martínez García (2018) found that bilinguals took less time to name Spanish targets with stress-sharing cognate competitors than with no-stress-sharing cognate competitors. Martínez García (2018, p. 20) interpreted this finding as evidence that "English stress modulates cross-language activation during bilingual spoken word production". It is important to note that, in this study, the bilinguals' two languages, namely English and Spanish, both have lexical stress, and only near-identical cognates were examined. Given that cognates might share "one single memory token" in bilinguals' minds, the observed stress effect may not be due to cross-language co-activation but result from the shared stress representation (or one stress assigning rule) in an integrated bilingual lexicon (Roelofs, 2003; 2006). Overall, it is still unclear whether suprasegmental information is co-activated during spoken word production, especially for bilinguals whose two languages have different systems of suprasegmental contrasts.

In this study, we aimed to fill this knowledge gap by investigating the activation of suprasegmental information with bilinguals of SC and English. SC is a representative tonal language, which uses pitch variation to differentiate morpheme or word meanings, just as consonants and vowels. For example, *ma* means "mother" when it is produced with a level pitch contour, but "horse" with a low-dipping pitch contour. Unlike tonal languages, stress languages such as English employ relative prominence between syllables to distinguish words (e.g., REcord and reCORD), which is cued not only with salient pitch contours but also salient lengthening, intensity increase, and vowel quality contrast (see Gordon & Roettger, 2017 for a review on cues of stress). Furthermore, unlike tonal contrast, which is abundant in Mandarin, the number of stress minimal pairs in English is limited (Giegerich, 1992). In short, the way suprasegmental information is utilized differs significantly in the lexicons of SC and English, which offers a unique case for investigating the interplay of phonological representations and language co-activation at the suprasegmental level.

We employed the PWI paradigm and asked SC-English bilinguals to name pictures in English while ignoring simultaneously presented SC distractors.

Critically, we manipulated the phonological overlap between the target's SC translation and the distractors. As shown in Table 1, for the same target (e.g., feather), there are four types of distractors: 1) translation distractor, which is the target's SC translation (e.g., *yu3mao2* "feather"); 2) tone-sharing (phono-translation) distractor, which shares both segments and lexical tone with the target's SC translation in the first syllable (e.g., *yu3zhou4*, "universe"); 3) no-tone-sharing (phono-translation) distractor, which shares only segments with the target's SC translation in the first syllable (e.g., *yu4mi3*, "corn"); 4) unrelated distractor, which has neither segmental nor tonal overlap with the target's SC translation (e.g., *lei4shui3*, "tear"). All the target and distractor pairs are not semantically related within and across languages. If lexical tone plays an important role in cross-language activation, we expect to find a significant difference between naming latencies in the tone-sharing and no-tone-sharing conditions.

Table 1. *A set of sample stimuli. The SC syllables are spelt out in Pinyin, an alphabetic writing system of SC. The numbers in Pinyin here indicate the lexical tone.*

	Target	Distractors			
		Translation	*Tone-sharing*	*No-tone-sharing*	*Unrelated*
English	feather	feather	universe	corn	tear
Pinyin	yu3mao2	yu3mao2	yu3zhou4	yu4mi3	lei4shui3
Character	羽毛	羽毛	宇宙	玉米	泪水

Moreover, we manipulated two procedural factors in the PWI. One is the modality of the distractors, namely whether participants listened to or viewed distractor words during picture naming (auditory vs. visual distractors). The other is the familiarization mode, 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) of the pictures before naming pictures in English.[7] These manipulations were designed to clarify the following specific issues in our understanding of language co-activation during bilingual word production.

The first issue concerns the different inhibition or facilitation effects with different distractor modalities. Despite that both *translation facilitation* and *phono-translation interference* effects have been used as indicators for language co-activation, it has been controversial how to reconcile the fact that the two effects are opposite (see Costa, 2005 and Hall, 2011 for reviews on this issue). Specifically, why would the phono-translation distractor, which is phonologically related to the targets' translation, inhibit target picture naming, while the translation distractor itself facilitates picture naming? Based on these observations, Costa (1999; 2003) proposed a language-specific lexical selection theory: words of the non-target language are excluded from lexical selection so that co-activated translations cause no lexical interference but semantic facilitation to target naming. By contrast, Hermans and his colleagues (Hermans et al., 1998; Hermans, 2004) argued for a language-nonspecific lexical selection account: there are two mechanisms underlying the translation and phono-translation effects, namely semantic facilitation at the conceptual level and cross-language competition at the lexical level; for translation distractors, semantic facilitation overrides lexical competition, leading to a net priming effect; for phono-translation distractors, the relatively weak semantic facilitation cannot overrule lexical competition, resulting in an interference effect. Despite the two opposing views, it is important to note that previous studies have mainly observed the translation facilitation effect with visual distractors (e.g., Costa et al., 1999; Costa & Caramazza, 1999; Hermans, 2004) and the phono-translation interference effect with auditory distractors (e.g., Hermans et al., 1998; Costa et al., 2003). Given that within-

---

[7] Typically, in PWI, there is a familiarization session before performing the naming task which allows participants to preview the target pictures and become familiarized with the intended target names.

language semantic distractors were typically found to be facilitative in the visual modality but inhibitive in the auditory modality (e.g., Hantsch et al., 2009; Jonen et al., 2021), the "translation and phono-translation paradox" (Hall, 2011) may simply be an artefact of distractor modality. To test this possibility and therefore gain a better understanding on the effect of lexical tone activation, in this study, we examined the effects of translation and phono-translation distractors in both visual and auditory domains. If indeed the contrast is caused by modality difference, we expect to replicate the translation facilitation effect only with visual distractors and the phono-translation interference effect only with auditory distractors. However, if the two effects are indeed opposite despite the modality difference, the central controversy may lie in whether language co-activation necessarily causes competition in language selection (i.e., the view of language-specific lexical selection vs. the view of language-nonspecific lexical selection; Hermans et al., 1998; Costa et al., 2003).

The second related issue concerns the conflicting views of language-specific and non-specific lexical selection, which we aimed to shed light upon by examining the consequences of increasing the activation level of the non-target language. If co-activation leads to cross-language competition as indicated by the language non-specific lexical selection view (Hermans et al., 1999), higher activation of the non-target language may interfere with target selection and prolong the naming latency. To this end, we further manipulated the SC activation level by adjusting the familiarization mode. Specifically, we expect to introduce a higher SC activation level by providing both English and SC target names in the familiarization session (mixed mode) than English names only (English mode). If we find a larger interference effect in the mixed mode over the English mode, this suggests the involvement of cross-language competition, lending support to the language non-specific lexical selection view (Hermans et al., 1999).

In sum, we aimed to answer whether SC lexical tone is co-activated during English picture naming by investigating the translation and phono-translation effects 1) across different distractor modalities and 2) with familiarization modes. By doing so, we hope to gain a more comprehensive understanding of the interaction between bilinguals' two languages. More specifically, we conducted four PWI experiments with four types of distractors (i.e., translation, tone-sharing distractor, no-tone-sharing, and unrelated distractors). In Experiment 1, participants were familiarized with targets' English names only and then performed picture naming with the presence of auditory distractors. In Experiment 2, participants were familiarized with the targets' English and SC names and then performed naming tasks with auditory distractors. Experiments 3 and 4 are replications of Experiments 1 and 2, respectively, using visual distractors.

4.1 Experiment 1 (Auditory Distractor and English Mode)

4.1.1 Method

4.1.1.1 Participants

Forty-one SC-English bilinguals (30 females and 11 males; average age 24) participated in this experiment. All participants are native SC speakers who grew up in Northern China. They speak no local dialect and have no history of language disorder. All participants passed College English Test Band 6 or scored above 6 in International English Language Testing System (IELTS). We also assessed participants' English proficiency level with an adapted LEAP-Q questionnaire (Marian et al., 2007) and the multilingual naming test (MINT; Gollan et al., 2012) which has been validated by several studies (e.g., Sheng et al., 2014 with Chinese-English bilinguals). All participants learned English at an average age of 8.0 (SD = 3.0). Using a Likert scale from 1 to 10, the mean self-rated proficiency by the participants was 8.0 (SD = 1.6) in reading, 6.7 (SD = 1.7) in speaking and 7.0 (SD = 1.7) in listening. The average correct response of MINT was 58% (SD = 11%). This study was approved by the Ethics Committee of the

Faculty of Humanities at Leiden University. All participants provided informed consent and were compensated 30 RMB for their participation.

4.1.1.2 Stimuli

There are 24 sets of critical stimuli (see Appendix C). Each set consists of an English target word, an SC translation distractor, an SC tone-sharing distractor, an SC no-tone-sharing distractor, and an SC unrelated distractor. There are also 12 sets of filler words which are not phonologically or semantically related. All words are common disyllabic nouns. Word frequency of SC and English, as computed with SUBTLEX-CH (Cai & Brysbaert, 2010) and SUBTLEX-US (Brysbaert & New, 2009), are balanced across conditions [SC: $F(3, 92) = 1.97$, $p = 0.13$; English: $F(3, 92) = 1.76$, $p = 0.16$]. Word length in English was also controlled [$F(3, 92) = 0.753$, $p = 0.52$]. The target pictures, which are black and white line drawings, were selected from the IPNP database (Bates et al., 2003) and the BOSStimuli database (Brodeur et al., 2012). Twenty-seven native Mandarin speakers who did not participate in the PWI experiments validated the choices of target stimuli in terms of picture naming agreement, translation agreement, and picture imageability of distractors. All spoken stimuli were recorded by a female native SC speaker (age 22) who was born and grew up in Beijing. The recording was done at the Leiden University Centre for Linguistics Phonetics Lab through a Sennheiser MKH416T microphone (44.1 kHz, 16 bit). All stimuli were normalized for duration of 1,000 ms and intensity at 70 dB using Praat (Boersma & Weenink, 2022).

4.1.1.3 Procedure

Participants performed the experiment online using Gorilla (www.gorilla.sc). All participants were required to wear headphones and sit in a quiet room. Participants were only allowed to join the experiment if using laptops. Prior to the experiment, a headphone task based on the dichotic pitch (Milne et al., 2020), as well as a microphone check and an auto-play check were run to screen participants' equipment. All the instructions were given in English. Before the picture naming task, there was a familiarization session. During this session, participants were shown 36 target pictures (24 critical and 12 filler targets) with their matching English names printed underneath. Afterwards, participants were asked to type in the picture names in English. If participants did not respond accurately, the intended name would appear again. In the PWI task, a fixation was displayed in the centre of the screen for 500 ms, followed by a picture and simultaneously played English spoken distractors (Stimulus Onset Asynchrony = 0 ms). Participants were asked to name the picture as quickly and accurately as possible while ignoring the auditory distractor. The picture remained visible for 2,000 ms. Response times were measured from picture onset until naming onset using Chronset (Roux et al., 2017). If participants did not respond within 2,000 ms, the trial ended, and the experiment proceeded automatically. Between each trial, there was a blank screen of 1,000 ms. Before starting the task, participants were asked to complete four practice trials, with an option to practice more rounds. In total, there were 96 (24 targets × 4 conditions) critical trials and 48 (12 targets × 4 conditions) filler trials. All the trials were equally distributed into four blocks in a Latin Square design so that participants only saw each target picture once in every block. Between each block, participants were encouraged to take a short break without changing the equipment set-up. After the PWI task, participants were asked to complete the MINT test and a language background survey. In total, the experiment took about 30 minutes.

4.1.2 Data Analysis

Response times (hereafter RT) were analysed using the generalized linear mixed-effects model (GLMM) with inverse Gaussian distribution (Lo & Andrews, 2015). Incorrect responses (e.g., responses in SC), blank responses and unrecognizable responses were excluded from data analysis. For each experiment, a maximum model including fixed effects of distractor type (i.e., translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors), random slope for distractor type by subject and item, and random intercepts for subject and item were constructed first. If a model failed to converge, we increased the number of iterations and then simplified the model by removing correlation parameters in the random structures (Brauer & Curtin, 2018). All the analyses were run in R Studio (R Core Team, 2022) with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). Pairwise comparisons were computed using the *multcomp* package (Hothorn et al., 2022). Holm–Bonferroni method was implemented to correct family-wise errors (Holm, 1979).

4.1.3 Results

Incorrect trials (~2.9%), trials with no response (~2.9%), and unrecognizable responses (~0.2%) were excluded from the analysis. Given that error rates were low in each condition, no further analysis on accuracy was conducted. Table 2 and Figure 1 summarise the mean RT for each condition. Compared with unrelated distractors, participants took about 29 ms longer to name pictures with tone-sharing distractors (i.e., tone-sharing condition), about 10 ms longer with no-tone-sharing distractors (i.e., no-tone-sharing condition), and about 28 ms less with translation distractors (i.e., translation condition). Moreover, response time in the tone-sharing condition was about 19 ms longer than in the no-tone-sharing condition. The final GLMM consists of the fixed effects of distractor type (i.e., translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors), random slope for distractor type by subject and item, and random intercepts for subject and item. According to the GLMM estimations (see Table 3), the interference effect in the tone-sharing condition ($p < 0.01$) and the facilitation effect in the translation condition ($p < 0.05$) are both statistically significant. Crucially, naming latency in the tone-sharing condition is significantly longer than in the no-tone-sharing condition ($p < 0.05$).

Table 2. *Mean naming latency and standard deviations of Experiment 1 (Auditory Modality and English Mode), Experiment 2 (Auditory Modality and Mixed Mode), Experiment 3 (Visual Modality and English Mode), and Experiment 4 (Visual Modality and Mixed Mode). The mixed effect shows the naming latency difference between the Mixed Mode and the English Mode (Mixed – English).*

Modality	Auditory Modality					Visual Modality				
Familiarization mode	English		Mixed		Mixed Effect	English		Mixed		Mixed Effect
Conditions	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Tone-sharing	1001	367	1064	348	64	1039	375	1002	290	-37
No-tone-sharing	982	360	1058	342	76	1033	319	1012	294	-21
Translation	944	329	1059	342	115	908	296	899	248	-9
Unrelated	972	351	1043	329	71	1044	348	1035	323	-8
Tone – Unrelated	29		21		-7	-5		-34		-29
No-tone – Unrelated	10		15		5	-11		-23		-12
Tone – No-tone	19		6		-12	6		-11		-17
Translation – Unrelated	-28		16		44	-136		-137		0

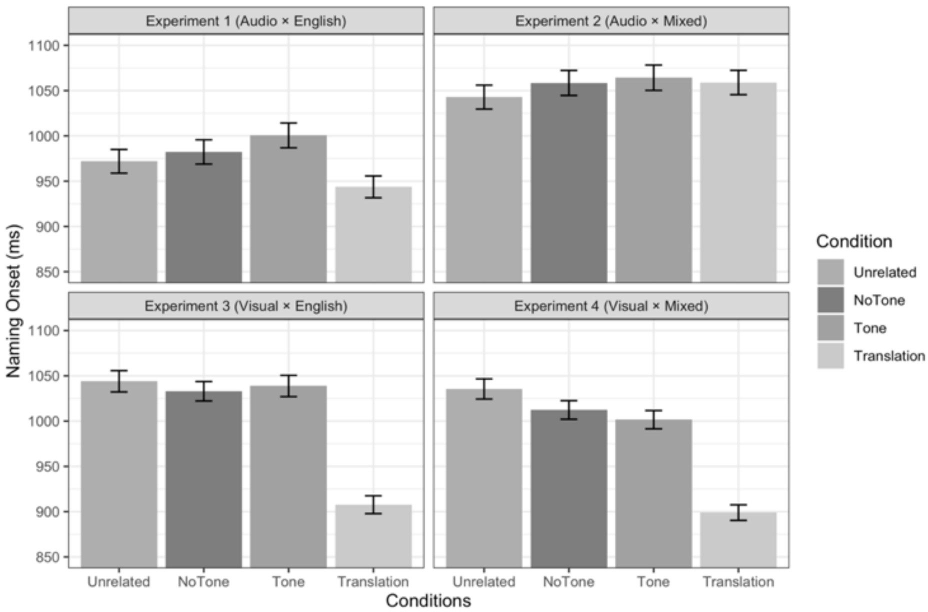

Figure 1. *Mean naming latencies of all conditions in Experiment 1, Experiment 2, Experiment 3, and Experiment 4.*

Table 3. *GLMM analysis of naming latency in Experiment 1.*

	Estimate	SE	t	*p*-value
(Intercept)	1160.440	19.200	60.440	<0.001
No-tone Sharing – Unrelated	14.540	13.070	1.112	0.266
Tone Sharing – Unrelated	54.760	15.950	3.433	0.002
Translation – Unrelated	-41.040	14.730	-2.786	0.016
No-tone Sharing – Tone Sharing	40.230	16.370	2.458	0.028

4.1.4 Discussion

In this experiment, we found that, compared with unrelated auditory distractors, target picture naming was significantly speeded up by SC translation distractors but slowed down by tone-sharing distractors. Such findings replicated previously found *translation facilitation* (e.g., Costa et al., 1999, 2003; Hermans,

2004) and *phono-translation interference* effect (e.g., Costa et al., 2003; Hermans et al., 1998) with auditory distractors. Importantly, there was a significant naming latency difference between the tone-sharing and no-tone-sharing conditions, indicating that lexical tone is co-activated and plays a significant role during the process of English picture naming.

4.2 Experiment 2 (Auditory Distractor and Mixed Mode)

4.2.1 Method

4.2.1.1 Participants

Forty-two SC-English bilinguals (31 females and 11 males; average age 24) participated in this experiment. All participants learned English at an average age of 8.6 (SD = 3.1). The mean self-rated frequency by participants was 8.1 (SD = 1.4) in reading, 6.8 (SD = 1.4) in speaking and 7.2 (SD = 1.7) in listening. The average correct response of MINT was 61% (SD = 10%).

4.2.1.2 Stimuli, Procedure & Data analysis

During the familiarization session, participants were provided with both English and SC names (i.e., mixed mode). By doing so, we expect to introduce a higher level of SC activation. If we find larger interference effects across distractors in Experiment 2 than Experiment 1, this suggests the involvement of cross-language competition in bilingual picture naming. Besides this, the same stimuli, procedure and analysis as in Experiment 1 were used in this experiment.

4.2.2 Results

Incorrect trials (~3.1%), trials with no response (~1.6%), and unrecognizable responses (~0.1%) were excluded from the analysis. Given that error rates were low in each condition, no further analysis on accuracy was conducted. Table 2 and Figure 1 summarize the mean RT for each condition. Compared with the unrelated condition, participants took about 21 ms longer to name pictures in the tone-sharing condition and about 15 ms longer in the no-tone-sharing and translation condition. The final GLMM consists of the fixed effects of distractor type (i.e., translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractor), random slope for distractor type by subject and item, and random intercepts for subject and item. According to model estimations (see Table 4), although the tone-sharing, no-tone-sharing and translation condition all exhibited inhibitory trends towards picture naming, none of them significantly differed from the unrelated condition (tone-sharing condition: $p = 0.227$; no-tone-sharing condition: $p = 0.547$; translation condition: $p = 0.547$).

Table 4. *GLMM analysis of naming latency in Experiment 2.*

	Estimate	SE	t	*p*-value
(Intercept)	1211.350	21.880	55.363	<0.001
No-tone Sharing – Unrelated	10.150	15.640	0.649	0.547
Tone Sharing – Unrelated	33.090	17.370	1.905	0.227
Translation – Unrelated	23.150	19.350	1.196	0.547
No-tone Sharing – Tone Sharing	22.940	17.200	1.334	0.547

As shown in Table 2, naming latencies in Experiment 2 are longer than in Experiment 1 across conditions. To further investigate the effect of familiarizing both English and SC names, a joint analysis of Experiment 1 and 2 was conducted. The final joint GLMM includes the fixed effect of distractor type (i.e., translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors), familiarization mode (English Mode in Experiment 1 vs. Mixed Mode in Experiment 2), the interaction between distractor type and familiarization mode, random intercepts for subjects and items, by-subject random slope for distractor type, by-item random slope for distractor type and familiarization mode, and their interaction. Results of the analysis showed a significant main effect of familiarization mode ($p < 0.001$) and a significant interaction between condition and familiarization mode ($p < 0.05$). Pairwise comparisons showed that naming latency of all conditions in Experiment 2 were significantly longer than that of Experiment 1 (Translation: $p < 0.001$; Tone-sharing: $p < 0.05$; No-tone-sharing: $p < 0.05$; Unrelated: $p < 0.05$).

4.2.3 Discussion

When bilingual participants were familiarized with both English and SC names (instead of English names only as in Experiment 1), we failed to replicate the previously found *translation facilitation* (e.g., Costa et al., 1999, 2003; Hermans, 2004) and *phono-translation interference* effect (e.g., Hermans et al., 1998; Costa et al., 2003) with auditory distractors. None of the distractors (i.e., translation, tone-sharing and no-tone-sharing) in Experiment 2 had a significant impact on the picture naming latency compared with unrelated distractors. Importantly, a joint analysis showed that the naming latency of all conditions was significantly longer in Experiment 2 (mixed mode) than in Experiment 1 (English mode). This indicates that introducing SC names alongside English names increases the processing demands involved in English picture naming.

4.3 Experiment 3 (Visual Distractor and English Mode)

4.3.1 Method

4.3.1.1 Participants

Forty-three SC-English bilinguals (34 females and 9 males; average age 24) participated in this experiment. All participants learned English at an average age of 7.3 (SD = 2.8). The mean self-rated frequency by participants was 8.1 (SD = 1.6) in reading, 7.0 (SD = 1.8) in speaking and 7.5 (SD = 1.8) in listening. The average correct response of MINT was 61% (SD = 13%).

4.3.1.2 Stimuli, Procedure & Data analysis

Experiment 3 is a counterpart of Experiment 1. Instead of presenting auditory distractors, the SC distractor words were superimposed on the target pictures as Chinese characters.

4.3.2 Results

Incorrect trials (~3.3%), trials with no response (~3.4%), and unrecognizable responses (~0.1%) were excluded from the analysis. Given that error rates were low in each condition, no further analysis on accuracy was conducted. Compared with unrelated distractors, participants took 136 ms less to name pictures when translation distractors were present, and 5 ms and 11 ms less, respectively, with tone-sharing and no-tone-sharing distractors (see Table 2 and Figure 1). The final GLMM consists of the fixed effects of distractor type, by-subject and by-item random slope for distractor type, and random intercepts for subject and item. According to the GLMM estimation (see Table 5), only the naming latency in the translation condition was significantly different from the unrelated condition ($p < 0.001$), showing a robust translation facilitation effect. Although both tone-sharing and no-tone-sharing conditions exhibited a small facilitatory trend toward picture naming, the response times did not significantly differ from the unrelated condition (tone-sharing condition: $p = 0.631$; no-tone-sharing condition: $p = 0.451$).

Table 5. *GLMM analysis of naming latency in Experiment 3.*

	Estimate	SE	t	*p*-value	
(Intercept)	1195.080	33.051	36.159	<0.001	***
No-tone Sharing – Unrelated	-17.542	12.728	-1.378	0.451	
Tone Sharing – Unrelated	6.384	13.279	0.481	0.631	
Translation – Unrelated	-161.933	19.143	-8.459	<0.001	***
No-tone Sharing – Tone Sharing	23.926	16.638	1.438	0.451	

Given that Experiment 3 is a replication of Experiment 1 with visual distractors, a joint analysis of Experiment 3 and 1 was conducted to investigate the impact of distractor modality. The final joint GLMM includes the fixed effect of distractor type (translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors), distractor modality (Auditory Modality in Experiment 1 vs. Visual Modality in Experiment 3), random intercepts for subjects and items, by condition random slope for subjects, by condition and by modality random slope for items. Results of GLMM analysis showed a significant main effect of distractor modality (p = 0.010) and a significant interaction between condition and modality (p < 0.001). Although naming latencies of tone-sharing, no-tone-sharing, and unrelated conditions in Experiment 3 were longer than Experiment 1, the differences were not statistically significant (tone-sharing: p = 0.071; no-tone-sharing: p = 0.190; unrelated: p = 0.190). However, the naming latency of the translation condition was significantly shorter in Experiment 3 than in Experiment 1 (p < 0.001).

4.3.3 Discussion

With visual distractors, we found robust *translation facilitation* effect, replicating findings from previous studies (e.g., Costa et al., 1999, 2003; Hermans, 2004) and from Experiment 1 with auditory distractors. However, in contrast to the previously found *phono-translation inhibition* effect, the tone-sharing and no-tone-sharing (phono-translation) distractors did not exhibit inhibition but rather insignificant facilitatory trends toward English picture naming.

4.4 Experiment 4 (Visual Distractor and English Mode)

4.4.1 Method

4.4.1.1 Participants

Thirty-eight SC-English bilinguals (31 females and 7 males; average age 24; SD = 1.6) participated in this experiment. All participants learned English at an average age of 7.3 (SD = 2.9). The mean self-rated proficiency by participants was 8.3 (SD = 1.6) in reading, 7.0 (SD = 2.1) in speaking, and 7.4 (SD = 2.0) in listening. The average correct response of MINT was 61% (SD = 13%).

4.4.1.2 Stimuli, Procedure & Data analysis

Experiment 4 is a counterpart of Experiment 2 with visual distractors. During the familiarization session, participants were provided with both English and SC names when familiarizing themselves with the target pictures.

4.4.2 Results

Incorrect trials (~1.8%), trials with no response (~1.7%), and unrecognizable responses (~0.1%) were excluded from the analysis. Given that error rates were low in each condition, no further analysis on accuracy was conducted. Table 2 and Figure 1 summarise the mean RTs for each condition. Compared with unrelated distractors, participants took about 137 ms less to name pictures when translation distractors were present; 34 ms less with tone-sharing distractors; 23 ms less with no-tone-sharing distractors. The final converged GLMM consists of the fixed effects of distractor type (translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors) and random intercepts for subject and item. According to the GLMM estimation (see Table 6), naming latency in the tone-sharing condition is significantly different than in the unrelated condition ($p < 0.05$); while the no-tone-sharing condition is not ($p =$

0.359). Moreover, with an estimated difference of about 117 ms, the translation condition is significantly different from the unrelated condition, showing a robust *translation facilitation* effect.

As we can see from Table 2, naming latencies in Experiment 4 were shorter than in Experiment 3 across all conditions. To further examine the effect of mixed familiarization mode in the visual modality, a joint analysis of Experiments 3 and 4 was conducted. The final joint GLMM includes the fixed effect of distractor type (translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors), familiarization mode (English Mode in Experiment 3 vs. Mixed Mode in Experiment 4), and random intercepts for subjects and items. Results of GLMM analysis showed a significant interaction between distractor type and familiarization mode ($p = 0.044$), and no significant main effect of familiarization mode was found ($p = 0.748$). Pairwise comparisons showed that there was no significant difference between Experiment 3 and 4 across conditions (Translation: $p = 0.198$; Tone-sharing: $p = 1.000$; No-tone-sharing: $p = 1.000$; Unrelated: $p = 0.832$).

Given that Experiment 4 is a replication of Experiment 2 with visual distractors, a joint analysis of Experiments 4 and 2 was also run to test the impact of distractor modality. The final joint GLMM includes the fixed effect of conditions, modality (auditory modality in Experiment 2 vs. visual modality in Experiment 4), random intercepts for subjects and items, by-subject random slope for distractor type, by-item random slope for distractor type and for modality. Results of the GLMM analysis showed a significant main effect of distractor modality ($p < 0.001$) and a significant interaction between condition and modality ($p < 0.001$). Pairwise comparison showed significant differences between Experiments 4 and 2 in all conditions (Translation: $p < 0.001$; Tone-sharing: $p < 0.001$; No-tone-sharing: $p < 0.001$; Unrelated: $p < 0.005$).

Table 6. *GLMM analysis of naming latency in Experiment 4.*

	Estimate	SE	t	*p*-value
(Intercept)	1133.721	22.673	50.003	<0.001
No-tone Sharing – Unrelated	-10.598	10.062	-1.053	0.359
Tone Sharing – Unrelated	-24.632	9.398	-2.621	0.026
Translation – Unrelated	-117.989	8.965	-13.161	<0.001
No-tone Sharing – Tone Sharing	-14.033	10.454	-1.342	0.359

4.4.3 Discussion

With visual distractors, the *translation facilitation* effect found in previous studies (e.g., Costa et al., 1999, 2003; Hermans, 2004) was replicated, same as in Experiments 1 and 3. The *phono-translation interference* effect (e.g., Hermans et al., 1998; Costa et al., 2003) in either tone-sharing or no-tone-sharing distractors were not found, unlike in Experiment 1. However, tone-sharing distractors introduced a significant facilitation effect to English picture naming whereas no-tone-sharing distractors did not. This indicates a crucial role of lexical tone in the process of bilingual spoken word production. Moreover, with visual distractors, familiarizing with target names in both languages or in English only, had no significant impact on the naming latency. Furthermore, with the mixed mode, auditory distractors interfered with the naming process significantly more than visual distractors.

4.5 General Discussion

To investigate the role of lexical tone in bilingual word production, four PWI experiments were conducted. SC-English bilinguals were instructed to name pictures in English while ignoring simultaneously presented SC distractors. Together with a target picture (e.g., “feather”), there were four types of SC distractors: a translation distractor (e.g., *yu3mao2* “feather”), a tone-sharing (phono-translation) distractor (e.g., *yu3zhou4* “universe”), which shares both segments and tone with the target’s translation in the first syllable, a no-tone-sharing (phono-translation) distractor (e.g., *yu4mi3* "corn"), which shares only segments with the target's translation in the first syllable, and an unrelated distractor (e.g., *lei4shui3* "tear"). Moreover, to gain a fuller understanding of the effects of co-activation, we manipulated the four experiments along two dimensions. One was distractor modality, i.e., the SC distractors were presented auditorily in Experiments 1 and 2, but visually in Experiments 3 and 4. The other was familiarization mode. Bilinguals were familiarised with English names only in Experiments 1 and 3 (English mode) but with both English and SC names in Experiments 2 and 4 (mixed mode). In Experiment 1 (auditory distractor and English mode), compared with unrelated distractors, we found significantly shorter naming latency with translation distractors and significantly longer naming latency with tone-sharing distractors (but not with no-tone-sharing distractors), replicating the *translation facilitation* and *phono-translation interference* effects found in previous studies (e.g., Costa et al., 1999, 2003; Hermans, 2004). 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), naming latencies across conditions were significantly longer than those in Experiment 1 (auditory distractor and English 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 strong 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 strong 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.

For the first time in the literature, our findings show that SC lexical tone is co-activated during English spoken word production. Moreover, we found that the previously found *translation facilitation* and *phono-translation interference* effects are not fixed but rather dynamic, depending on procedural factors such as the distractor modality and familiarization mode: both *translation facilitation* and tone-sharing *phono-translation interference* effects were replicated using auditory distractors; visual distractors significantly strengthened *translation facilitation* effects and switched the tone-sharing *phono-translation interference* into facilitation. Moreover, mixing two languages during the familiarization session elicited more facilitation with visual distractors, but more interference with auditory distractors. Implications of these findings are discussed in the following sections.

4.5.1 The Role of Lexical Tone in Speech Production

In Experiment 1, we found a significant naming latency difference between tone-sharing and no-tone-sharing conditions. This suggests when bilinguals were naming target pictures in English, the lexical tone of the SC translations was also activated. This finding not only contributes to our understanding of language co-activation but also provides implications for tone word production models.

According to one of the most dominant models of speech production, i.e., the WEAVER++ model (Jescheniak & Levelt, 1994; Roelofs, 2000, 2015), to successfully generate a spoken word in stress languages such as English, speakers need to retrieve both phonological content and metrical frame during phonological encoding. While phonological content includes a set of ordered segments (phonemes), the metrical frame consists of syllabic information (e.g., number of syllables) and suprasegmental information (e.g., stress patterns). After retrieval, segments and the metrical frame associate and form a phonological word, i.e., a sequence of syllable(s). To account for tone word production, Roelofs (2015) proposed a level of tonal frame which functions similarly to the metrical frame. However, recent findings of tone error seem to suggest a more critical role of lexical tone. In a large Cantonese natural conversation corpus, Alderete et al.

(2019) found that tone errors are common (over 20% of the total sound errors) and tend to be influenced by adjacent tones in the same fashion as segmental errors. These findings prompt Alderete and his colleagues to propose that lexical tone is independently represented and equally selected as segments. Our finding of lexical tone co-activation in bilingual spoken word production seems to add more evidence to the view of Alderete et al. (2019). If lexical tone is represented diacritically as a tonal frame and not actively selected until phonetic spell-out, we are unlikely to observe any effect of lexical tone during bilingual word production of a non-tonal language. Together with the findings of previous studies, our results indicate that speakers of tonal languages are likely to select and encode lexical tone just as segments regardless of whether speaking in their native tonal language or the non-tonal second language.

4.5.2 The Translation Facilitation and Phono-Translation Interference Effects

As discussed in the introduction, previous studies have not reached a consensus on how to reconcile the seemingly contrasting *translation facilitation* and *phono-translation interference* effects (Costa, 2005; Hall, 2011). One possibility is that the contrast is introduced using different distractor modalities. With auditory distractors, we replicated both *translation facilitation* and tone-sharing *phono-translation interference* effects in Experiment 1. It is thus unlikely that the opposite translation and phono-translation effects are artefacts of distractor modality. Instead, distractor modality has a significant impact on the magnitude and direction of the (phono-)translation effects. With visual distractors, we found stronger effects of *translation facilitation* than with auditory distractors. Moreover, with visual distractors, the auditory *phono-translation interference* effect turned into a facilitation effect. As discussed in Hantsch et al. (2009, p. 1451), who also observed opposite effects for auditory and visual modality, respectively, this may be “due to differences of the time course with which the semantic representation of the distractor becomes available.” Given the nature of parallel processing of visual distractors versus sequential processing of auditory distractors, semantic representations of visual distractors may become available more quickly than that of auditory distractors. As a result, facilitation at the semantic level introduced by visual (translation and phono-translation) distractors may be more likely to boost lemma activation and speed up target picture naming than auditory distractors.

As discussed earlier, there are two general views on the underlying mechanisms of *translation facilitation* and *phono-translation interference* effects. One is the language-specific selection view (Costa, 1999), which posits that co-activated translations facilitate target production at the conceptual level; the other is the language non-specific selection view (Hermans et al., 1998), which argues that co-activated translations not only introduce semantic facilitation but also interfere with the process of word selection. Our results on the mixed vs. English familiarization mode seem to agree with the latter view. As we increased the activation level of SC by exposing participants to both English and SC target names during the familiarization session, the picture naming latencies increased significantly across distractor types compared to when participants were presented with English names alone. Furthermore, the effect of *translation facilitation* was cancelled out.

However, it is important to note that the inhibitory effect of the mixed familiarization session was only found with auditory distractors (Experiment 1 vs. Experiment 2) but not with visual distractors (Experiment 3 vs. Experiment 4). There are two possible explanations for this divergence. First, with visual distractors, bilinguals viewed Chinese characters imposed on target pictures, which might have led them to encounter the ceiling level of cross-language interference; thus, further increasing SC activation by introducing SC names in the familiarization session did not elicit any significant inhibitory effect on target picture naming. Second, according to Hermans et al. (1998), increasing SC activation could result in not only more cross-language interference but also more semantic facilitation; given that visual distractors may have earlier access to semantic representations than their auditory counterparts, it is thus possible that the increased SC activation in mixed mode boost semantic facilitation more readily with visual distractors than auditory distractors and cancelled out part of the cross-language interference effect.

4.5.3 Methodological Contributions

The present study also has a few methodological contributions to the use of PWI in bilingual word production studies. First, this study validated the significant impact of distractor modality in PWI and extended findings of distractor modality's influence on monolingual semantic effects (e.g., Hantsch et al., 2009; Jonen et al., 2021) to bilingual translation and phono-translation effects. Second, we demonstrated that asking bilinguals to become familiarized with target pictures' names in both their languages is an effective way to adjust language activation levels, especially with auditory distractors. This could be an important factor to manipulate for future use of PWI in bilingual studies. Third, successfully replicating previous lab findings with online experiments, this study showed that virtual PWI is an efficient and sound approach to studying speech production.

4.6 Conclusion

In conclusion, the data reported in this study showed, for the first time, that lexical tone is co-activated during the process of bilingual spoken word production. Moreover, we found that the effect of co-activating translations and their lexical tone is greatly impacted by experimental details such as distractor modality (i.e., whether participants see or hear distractor words) and familiarization mode (i.e., whether participants are familiarized with picture names in the target language only or both the target and non-target languages before picture naming). These findings provide new insights for understanding language co-activation at the suprasegmental level and the role of lexical tone in spoken word production.