CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Cross-language interactions during concurrent comprehension and production: evidence from simultaneous interpreters

Xueni Zhang[1] 🄳, Binghan Zheng[1] 🄳 and Yan Jing Wu[2,3] 🄳

[1]School of Modern Languages and Cultures, Durham University, Durham, UK, [2]School of Linguistic Sciences and Arts, Jiangsu Normal University, Xuzhou, China and [3]Jiangsu Collaborative Innovation Center for Language Ability, Jiangsu Normal University, Xuzhou, China
Corresponding author: Binghan Zheng; Email: binghan.zheng@durham.ac.uk

Abstract

It has been established that bilinguals activate both languages even when only one language is being used. However, little is known about how the two languages are co-activated during simultaneous interpreting (SI), a demanding task involving intensive code-switching. This study investigated (1) the effect of task on cross-language co-activation and (2) the time course of co-activations triggered by form and meaning. Thirty-one professional interpreters were recruited to complete a cross-language task (English-to-Chinese SI) and a within-language task (English-to-English shadowing) with their eye movements tracked. Participants heard English passages which contained critical spoken words, each paired with a visual display of four Chinese words. One of the words was a competitor that resembled the translation equivalent of the spoken word in either form or meaning, and the other three were unrelated distractors. We found that participants directed more visual attention to both types of competitors at an early stage in shadowing, while the word-form competitor effect occurred during SI preceded that of the semantic competitor. Our findings support the parallel account of SI processing, with implications provided for the relationship between cross-language interactions and the time lag between input and output during interpreting.

Keywords: cross-language co-activation; eye movements; language non-selective lexical access; parallel processing; simultaneous interpreting

The ability to switch between two languages is a critical skill that bilinguals rely on. Some bilinguals can perform simultaneous interpreting (SI), which involves instantly rendering messages conveyed in one language into another under extreme time pressure. However, how such intensive and complex language operations differ from common bilingual processing is not yet fully understood. It is well established

that bilinguals' two languages are activated in parallel even when only one language is being used (e.g., Lagrou et al., 2011; Shook & Marian, 2019; Thierry & Wu, 2007; Weber & Cutler, 2004; Wu & Thierry, 2012b). In most bilingual settings, the two languages compete for activation. The unused language[1] must be suppressed in order to operate in situations where only one language is used exclusively. Such inhibition becomes even more complicated in a dual-language context in which one switches between languages in response to interlocutors' preferences. The two languages can also be used in a complementary manner when engaging in a community of dense code-switching bilingual speakers (Green & Abutalebi, 2013). In all these instances, however, one language is prioritized over the other at specific times to fulfill task goals. Activation of the unused language observed in these tasks is covert and often taken as a source of interference.

Different from regular bilingual tasks, SI entails the explicit use of both languages at the same time. It is intuitive to assume that cross-language parallel activations lead to translation naturally. However, SI in fact involves an even larger magnitude of cross-language competitions and interferences, requiring increased cognitive control (Dong & Li, 2019). Interpreters often face challenges due to the need to inhibit spontaneous activation or activation of less sufficient translation equivalents within a very short time span. Moreover, SI is often operated in an ongoing manner with comprehension and production overlapping each other, making disengagement from the non-target language more difficult. Do interpreters activate the target language unconsciously, as in the case of other bilingual tasks, and take advantage of that for production? Do they co-activate more strongly as a result of the intention to translate? In this study, we investigate cross-language co-activation of lexical items during concurrent comprehension and production in professional interpreters. Our specific focus concerns 1) whether co-activation in SI is different from that in shadowing (i.e., single-language verbatim repetition of the speech that is listened to simultaneously), and 2) whether co-activation is more likely to be triggered by similarity in phonology (and/or orthography) or by semantic relatedness.

Cross-language co-activation

Knowing two languages has a substantial effect on the cognitive mechanism underpinning language activation, as the same concept can be represented in completely different forms in the two languages. A key question is how bilinguals activate the language in use or inhibit the unused language to keep focused on the current language task? Research on cross-language interactions suggests that bilinguals' two languages are stored in a single, integrated lexicon. For instance, between-language cognates were recognized faster than noncognates both in visual (e.g., Dijkstra et al., 1999; Duyck et al., 2007; Van Hell & Dijkstra, 2002) and spoken word comprehension (Andras et al., 2022; Blumenfeld & Marian, 2007; Valente et al., 2018), suggesting parallel access of both languages involved in cognate processing (i.e., language non-selective access). Evidence also came from interlingual homographs or homophones, words that share certain aspects of their linguistic forms across languages (e.g., Chen et al., 2017; Dijkstra et al., 1999, 2000; Lagrou et al., 2011; Macizo et al., 2010). Lagrou et al. (2011), for instance, found that bilinguals responded significantly slower when hearing interlingual homophones

than control words irrespective of whether they listened to L1 or L2, supporting a strong account of language non-selectivity.

Other studies have shown that cross-language co-activation exists even in cases where the unused language was not elicited by providing a cross-language cue (i.e., an "All-in-L2 context"). Thierry and Wu (2007) recorded, by means of event-related potentials (ERP), the evidence that English words were unconsciously translated into Chinese counterparts when Chinese-English bilinguals performed a semantic relatedness task on English word pairs. In their study, the prime and target words were related either in meaning ("post" and "mail") or in the form of their Chinese translation equivalent ("train" as "火车" in Chinese and "ham" as "火腿"). The N400 effects, indicating spontaneous activation of Chinese (participants' native language), were found in both conditions, whereas participants reported being unaware of any information associated with the Chinese language. Such covert co-activation was also observed in Shook and Marian's (2019) study, in which English-Spanish bilinguals were asked in English to click on a picture showing a duck for example but looked more to a shovel because the word "duck" ("pato") and the word "shovel" ("pala") share the same phonological onset in Spanish.

The non-selective nature of bilingual processing, as consistently supported by empirical evidence (e.g., Jiang, 2021; Türker, 2018; Wu & Thierry, 2010, 2012a), is often followed by subsequent top-down, higher-level cognitive control (Dijkstra & van Heuven, 2002). However, views were more mixed regarding the extent to which top-down control modulates the degree of automatic co-activation. Evidence supporting the top-down modulation derives from the finding of target-language priority (priority of the language in use) during bilingual processing. In L2 sentence listening, FitzPatrick and Indefrey (2014) manipulated the semantic fit of sentence-final interlingual homophone in a sentence context as biased towards either the language in use or the unused language (target-biased and non-target-biased conditions) in addition to semantically fitting or unfitting sentences (fully congruent and fully incongruent conditions). The results showed an early onset of N400 in the non-target-biased condition compared to the fully incongruent condition and a negativity appeared later than the N400 in the target-biased condition, indicating that the two languages were activated sequentially with the language in use being prioritized. In a spoken word recognition study, Chen et al. (2017) found that the effect of semantic relatedness emerged approximately 500 ms after the onset of interlingual homophone, which was later than the within-language semantic effect observed in non-homophone control word pairs. The difference in terms of the time course of co-activation found in both studies showed that the unused language did not become accessible until after the meanings were processed in the language in use. It appears that language membership can impose restriction on non-selective access. However, such top-down modulation is not always observed. Lagrou et al. (2011) found that there was a limited role of cues provided by between-language subphonemic differences to restrict non-selective parallel access. This means that the fine-grained differences in the L1 and L2 pronunciations were not sufficient for participants to distinguish which language was in use. Van Hell and de Groot (2008) showed that the language in which the contextual information was

rendered did not restrict bottom-up activation of words in the unused language. Therefore, embedding words in a sentence does not necessarily provide a cue of language membership that modulates cross-language co-activation.

According to the BIA + model (Dijkstra & van Heuven, 2002), a task/decision system is involved in performing a bilingual task, which operates later in time and interacts with automatic, preconscious activations in the initial stage of processing. The task/decision system incorporates factors such as instructions, task demands, or participant expectancies, which does not impose direct influence on bottom-up processes but modulates them via the adaptation of decision criteria. It appears that the interaction between bottom-up co-activation and top-down modulation is sensitive to a number of factors such as the language mode (Grosjean, 1998), the nature of task, task modality, and task demand, and investigation into these factors might shed light on renewed interpretation for the mixed results observed by previous studies. Also, while cross-language co-activation has been investigated in a variety of linguistic levels and tasks, little is known about cross-language interactions when the current language in use is frequently switched between bilingual's L1 and L2.

The SI task, recruiting two languages in concurrent comprehension and production, is expected to generate strong parallel activation as switches are made between languages. To perform SI, a task schema that specifies the language-modality connections is required to be established. When interpreting from L2 to L1, for example, auditory input is connected to L2, while vocal output is linked to L1; the L1 and L2 connection is established to enable cross-language transposition. SI thus engages individuals in language-modality switches of predetermined frequency and intensity, with simultaneous involvement of two languages (Dong & Li, 2019). While regular bilinguals (i.e., those who are not trained to perform interpreting) suppress the activation of the unused language, interpreters are wired to instantly activate translations of the heard language.

Is such activation driven by non-selective lexical access? Previous research on interpreters' performance at prediction has shown that interpreting experience facilitates L2 morphological anticipation (Lozano-Argüelles et al., 2020; Lozano-Argüelles & Sagarra, 2021) and that semantic prediction takes place in SI regardless of training or experience (Amos et al., 2022; Liu et al., 2022). The observations that speeded lexical activation, or even pre-activation during interpreting, seem to suggest that the non-selectivity of bilingual access is being taken advantage by interpreters to achieve cross-language mapping and production in an efficient way. That being said, it is unclear whether such predicative behaviors observed are driven by pre-activations of upcoming information at the bottom-up, lower levels of representations (Kuperberg & Jaeger, 2016) or a reflection of some task-oriented, language comprehension strategies wielded by interpreters. Cross-language co-activations may also be absent in SI because interpreters have to exercise greater top-down control to coordinate between comprehension and production on the one hand and between the target language and the source language on the other hand. The heightened task demand might lead to more focused processing and thus reduced non-selective lexical access, a hypothesis supported by the findings from FitzPatrick and Indefrey (2014) and Chen et al. (2017).

The present study

The present study aims to explore the degree of cross-language co-activation during SI. A key feature of SI is the simultaneity of comprehension and production, which is one of the reasons why SI is more cognitively taxing compared to language tasks that involve either comprehension or production alone (Christoffels & de Groot, 2004). To control for this variable, we compare SI with shadowing. Like SI, shadowing requires comprehending and producing speech at the same time, but it involves only one language and may recruit less cognitive effort as no reformulation of message is needed. By comparing the two tasks, we will be able to differentiate cross-language production from single-language production, thereby revealing the co-activation effect. That being said, the results regarding the degree of semantic engagement during shadowing are mixed. On one hand, shadowers seem to follow the upcoming phonetic information verbatim without deeper levels of processing. In a comparison between picture naming and repetition in French, Gustafson et al. (2013) found that picture naming, which requires semantic processing, was more difficult and elicited slower, more accented speech than repetition. On the other hand, there is also evidence that shadowing indeed involves analysis up to a semantic level (Marslen-Wilson, 1973). We will take this into account when discussing findings regarding semantically driven co-activations.

SI in the present study is expected to be performed continuously, rather than on a sentence-by-sentence basis, as is in line with real-world scenarios. In this case, linguistic context information may modulate parallel activation to a certain degree (BIA+, Dijkstra & van Heuven, 2002). Chambers and Cooke (2009) demonstrated that the looks to near-homophone competitors reduced in semantically constraining sentences than control condition in a spoken word recognition task. However, when comparing cognate effects in low and high context conditions, van Hell and de Groot (2008) found that cognate effects remained after reading or translating in low-constraining sentence contexts but not in high-constraining conditions. Therefore, a more fine-grained finding is that only sentences with highly constraining context significantly confine non-selectivity. We thus decided to keep contextual information neutral for experimental sentences so that they are not constraining to show an effect of linguistic context, while global-level textual context was retained.

With shadowing as a baseline task, our first research question examines whether the extent of cross-language co-activation would be affected by the overt intention to translate between languages. There are three possible outcomes regarding task comparisons. First, if co-activation effects can be observed in SI and to an even greater extent than in shadowing, it would suggest that non-selective access is strengthened by the explicit intention to translate. It may also be that automatic activation of the target language helps facilitate conscious, deliberate processing of translation equivalents. Second, if co-activation effects can be observed in SI but to an indifferent extent as in shadowing, the magnitude of non-selective lexical access occurred in SI should be the same as that in other regular bilingual tasks. Third, if the co-activation effects are absent in SI, this could be a result of overly consumed cognitive resources by task, which lead to reduced bottom-up processing. Following

the BIA+ model, we predict that co-activations would be observed in both SI and shadowing, despite task nature. We also assume the results would be in line with the first possible outcome, given the evidence for fast lexical access during interpreting (Liu et al., 2022; Amos et al., 2022).

The second question of interest is how rapidly and to what extent co-activation can be triggered by form and meaning. We adopted a visual world eye-tracking approach to examining the time course of co-activation of L1 orthographic/ phonological and semantic information when listening and interpreting a passage in L2. In Paradis's (1994) conceptualization of SI, one strategy to perform translation is transcoding, i.e., direct mapping of translation equivalents between two languages. Another route of translation is to firstly decode the utterance in source language until conscious comprehension, which is then followed by linguistic encoding of message into the target language. In other words, translation is conceptually mediated, involving conscious semantic processing. This reflects, to some extent, a serial/sequential view of SI in which comprehension, transposition, and production take place separately (Seleskovitch, 1976). Based on studies showing cross-language interactions in single-language processing (Duyck et al., 2007; Lagrou et al., 2011; Thierry & Wu, 2007; Weber & Cutler, 2004; Wu & Thierry, 2010, 2012b), we predict that transcoding—i.e., the parallel account of SI—would serve as the primary mechanism underlying lexical activation processes in interpreters. If transcoding is adopted as we predict, we should be able to observe more rapid looks to words that share certain features of the word form of the translation equivalent of the heard word than those that are related in meaning. However, if SI follows the second strategy, the meaning of a word would be retrieved at an earlier stage than its corresponding lexical form in the target language, and participants should be more likely to fixate on L1 words that share meaning with those that they hear.

Given extensive literature on the language non-selective nature of lexical processing in bilinguals, it is reasonable to expect that direct mapping between translation equivalents in the two languages modulates SI processing. Paradis (1994) also argued that experienced interpreters are likely to exhibit transcoding considerably as the cross-language mappings are stored and retrieved more frequently as part of the task schema entailing SI performance. However, it should be noted that the top-down intention to speak proactively modulates lexical activation, which is largely meaning-based (Strijkers et al., 2011; Wu & Thierry, 2017). It may also be that the activation of a translation equivalent is mostly production-oriented and out of conscious intention, leaving little possibility for non-selective parallel activation. When involving concurrent operations of comprehension and production, it remains unclear whether conscious intention would dominate over bottom-up co-activation. We reason that activation of word form would appear earlier than semantic activation, but the subsequent semantic processing should attract a larger magnitude of visual attention as a result of proactive preparation for production. This prediction is consistent with the BIA+ model that early-stage activations are bottom-up and preconscious, followed by top-down control. Nonetheless, it should not be ruled out the possibility that interpreters can take advantage of the bottom-up route in which semantic processing is less involved.

Methods

Participants

Thirty-one English-Chinese professional interpreters (late fluent bilinguals) were recruited through snowball sampling. All of them were based in mainland China at the time the experiment was conducted. Data from 4 participants were excluded from analyses due to severe data loss as indicated by low gaze sample rates (<63.47%, participants' gaze sample mean minus one standard deviation). The final data consisted of 27 participants (5 males; *Mean age* = 31.33, *SD* = 5.88), with 7.11 years of interpreting experience on average (*SD* = 4.02). Mandarin Chinese was their native language, and they had learned English (L2) as a child (*Mean age of acquisition* = 8.48, *SD* = 3.24) through formal education. The mean length of their exposure to L2 is 22.85 years (*SD* = 5.54). On a 10-point scale of L2 proficiency, they self-rated as intermediate-to-advanced L2 user (*Mean L2 proficiency* = 7.96, *SD* = 0.94), and average frequency of L2 usage was rated as 4.48 (*SD* = 0.64) on a 5-point scale. All participants reported normal or corrected-to-normal vision. Informed written consent was obtained, and participants received monetary compensation for their participation. The study was approved by the Ethics Committee of Durham University.

Materials & stimuli

The experimental materials were composed of 6 spoken texts in English, each of which contained 18 trials, resulting in a total of 108 trials. Each trial consisted of a spoken word paired with a visual display (See Figure 1 for a sample trial). In the visual display, there were four two-character Chinese printed words: a critical word and three unrelated distractors. In the character repetition condition, the critical word shared the initial character with the Chinese translation equivalent of the spoken word (e.g., spoken word: *building*; translation equivalent: 建筑 *[Jian Zhu]*; competitor: 建议 *[Jian Yi—advice]*). In the semantic condition, the critical word was semantically related to the translation equivalent, but never overlapped phonologically or orthographically (e.g., spoken word: *agriculture*; translation equivalent: 农业 *[Nong Ye]*; competitor: 田地 *[Tian Di—farmland]*). In the filler condition, the critical word was the translation equivalent of the spoken word. The reason for including translation equivalents in filler trials was to engage participants' visual attention on screen throughout the experiment as they might be easily distracted from such cognitively taxing task as SI if visual information had little relevancy to what they were listening to.

Of each spoken text, the spoken words were embedded in individual sentences, distributed as evenly as possible across the text. While some of the spoken words appeared for multiple times in the text, it was ensured that only the first appearance of the words was included as trials. The sentences consisted of a mean of 15.89 words in the character repetition condition (*SD* = 7.02, *Range* = 4–30 words), 15.97 words in the semantic condition (*SD* = 6.46, *Range* = 6–28 words), and 15.89 words in the filler condition (*SD* = 7.25, *Range* = 5–32 words). The spoken words were at different positions in the sentences (character repetition condition: *Range* = 3–24th word, *Mean* = 11.25, *SD* = 5.83; semantic condition: *Range* = 6–28th word, *Mean* = 10.64, *SD* = 5.79; filler condition: *Range* = 4–28th word,
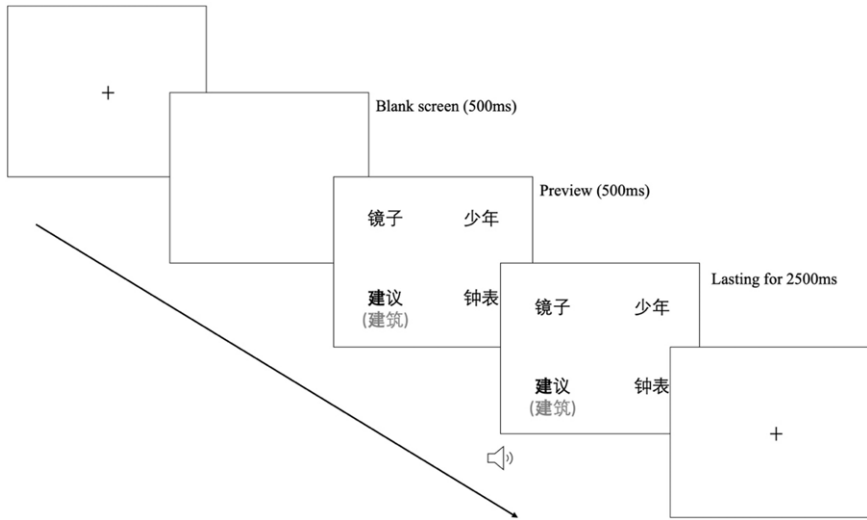
Figure 1. Sequence of trial events.
Notes: Visual onset preceded auditory onset of spoken word for 500 ms, following 500 ms of blank screen. Visual display remained on screen for 2500 ms since after auditory onset. This example shows a trial of character repetition condition. The spoken word *buildings* appeared in the utterance "And with the tourists, there has come all sorts of buildings." The translation equivalent of the spoken word was 建筑 *[Jian Zhu]*. The visual display consisted of the competitor: 建议 *[Jian Yi]*, and three unrelated distractors.

Mean = 13.08, *SD* = 5.87). None of the sentences contained more than one spoken word. Taken together, there were 6 trials of each condition in each text and therefore 36 trials in each condition in total (See Appendix 1 for a set of trials in a sample text). The presentation sequence of trials in each text was designed in a way that no more than 2 trials from the same condition occurred in direct succession.

Stimulus ratings
To assess validity of the stimuli, a pre-study was conducted. Fifteen Chinese-English bilinguals (1 male, *Mean age* = 24.3 years, *SD* = 1.5) who were studying in the UK and did not participate in the eye-tracking experiment completed a cloze probability test to help select experimental spoken words. Since we used texts as stimuli, predictability was measured in terms of both the local context provided by single sentences (lexical prediction) and the global context of the texts (graded prediction, Luke & Christianson, 2016). For each text, participants read the sentences truncated before potential experimental word candidates and completed the sentence fragments. They were instructed to use the first word that came to mind and were later asked to rate word predictability on a 5-point scale (1 representing *lowest predictability* and 5 representing *highest predictability*). Words that were easy to predict (i.e., cases where cloze probability is above .50) were excluded, so that concurrent activation, rather than pre-activation, could be captured. However, we do not assume the textual context in relation to the experimental words to be completely neutral, as this is seldom possible when listening to and comprehending a text. The mean predictability rating was 3.31

(SD = 0.76) for words in character repetition condition and 3.38 (SD = 0.71) for those in semantic condition. The participants also provided subjective ratings of word familiarity (5-point scale with 1 indicating *least familiar* and 5 indicating *most familiar*) on the stimulus words. We compared the mean word familiarity ratings for the two experimental conditions (character repetition: *Mean* = 4.91, SD = 0.10; semantic: *Mean* = 4.92, SD = 0.11).

The same group of participants performed a translation agreement task. They were asked to translate the experimental words into Chinese in a maximum of two Chinese characters. The words were embedded in sentences that were excerpted from the texts to provide context for comprehension. The translated words that were most frequently answered were taken as translation equivalents. We calculated word concreteness scores for the experimental words using word concreteness ratings developed by Brysbaert et al. (2013) and ensured that the words belonging to the two experimental conditions were comparable in terms of word concreteness (character repetition: *Mean* = 3.90, SD = 0.95; semantic: *Mean* = 3.95, SD = 0.83). We also controlled for age-of-acquisition (AoA) of the experimental words using the AoA ratings (Kuperman et al., 2012), as the order in which words were learned could potentially influence the speed and intensity of them being activated. The mean AoA rating was 5.99 (SD = 1.72) for words in character repetition condition and was 6.26 (SD = 2.01) for those in semantic condition.

Participants also rated the semantic relevancy for the translation equivalents of spoken words and competitors on a 7-point scale (1 representing *most relevant* and 7 representing *least relevant*). Word pairs in semantic condition (*Mean* = 2.34, SD = 0.39) were rated significantly more semantically relevant ($p < .000$) than words in character repetition condition (*Mean* = 5.67, SD = 0.72). We also made sure that distractors were semantically and logographically different from competitors as well as different from each other. Distractors that were rated relevant to competitors or other distractors in the same visual display by more than five participants were replaced. Additional rounds of ratings were administered until no relevancy was identified.

Textual properties
Six speeches were selected from the EU speech repository and transcribed verbatim. The transcripts were then trimmed into texts of comparable length but kept as well-structured and coherent as possible. Textual complexity was manipulated to make the texts comparable. Readability (i.e., Flesch Reading Ease) was chosen as a baseline indicator of text complexity. As SI involves both the grasp of global context and sentence-by-sentence transposition, a comprehensive array of measures at lexical, syntactic, and discourse levels was assessed using Coh-Metrix (Graesser et al., 2004) to capture subtle linguistic and textual variables. An additional measure was idea density. Sentences are comprised of meaning units, which are called ideas or propositions, and idea density is the ratio of the number of such units to the total words in a text. Idea density was calculated through propositional analysis using CPIDR 5 (Covington, 2009). See a summary of textual properties in Appendix 2 and Coh-Metrix parameters abbreviations and descriptions in Appendix 3. The texts

Table 1. Profile of spoken texts

Theme	Speed of delivery (in *words per minute*)	Duration	Word count
Children: asset or liability	105 wpm	05'30"	579
Light pollution	110 wpm	05'47"	634
Transport happiness	103 wpm	06'08"	633
Benefits of exercising	103 wpm	05'27"	563
Planning a holiday	110 wpm	05'49"	638
Electronic devices and waste	110 wpm	05'47"	633

were read aloud by a male native American English speaker and recorded via a microphone at a sample rate of 44.1 kHz. The speaker read the texts with a neutral intonation at a rate of mean 1.78 words per second. Table 1 shows the profile of the spoken texts.

Cued recall test

Cued recall test was used in this study to measure memory performance (Christoffels & de Groot, 2004). There were two aims for administering this measurement. First, the shadowing task is likely to trigger only surface-level semantic and syntactic processing. Even though semantic processing is found to be involved in shadowing as mentioned before, shadowers can repeat the speech literally with a grasp of the broad sense of the content. This is different from the case of SI, where sentence-by-sentence comprehension and reformulation are compulsory and thus speech input is processed much more profoundly. Shadowing with a post hoc recall task is expected to engage participants to a more conscious level, so that shadowing defined in the study involves comprehension in an explicit way. Second, we use cued recall test to compare the amount of working memory resources consumed by the two tasks. The task which requires larger consumption of working memory resources is expected to lead to reduced recall, as less resources would be left available for remembering stimuli (Christoffels & de Groot, 2004).

Only sentences embedding visual displays were selected as cued recall trials, so that the test would not be confounded by whether visual information was present or not. Of each spoken text, 9 sentences were selected as trial items, 3 containing spoken words of character repetition condition, 3 containing those of semantic condition, and 3 containing those of filler condition. Since the participants worked with 3 texts in each task, a total of 27 cued recall trials were completed in the two tasks respectively. For sentences that are too short to convey coherent messages, an additional sentence adjacent to it was also provided (See Appendix 4 for a sample test). In each sentence, a fragment, corresponding to 3.63 words on average (*Range* = 1–6) that constitutes a sensible meaning chunk, was presented at different positions of the sentence. Participants were required to recall and complete the sentence fragments drawing on available cues provided by the sentences as accurately as possible. The test was administered in a pen-and-paper format.

Procedure

Participants were seated 65 cm away from a viewing screen. Visual stimuli were presented on the screen at a resolution of 1920 × 1080 pixels, and spoken texts were presented to them through headphones. Eye movements were tracked using a Tobii Spectrum eye-tracker sampling at 600 Hz. The eye-tracker was calibrated using a five-point calibration method. Each spoken text began with a short bell sound as well as a drift correction appearing at the center of screen. Before a trial, there was a 500 ms blank screen. The visual display was then presented 500 ms before the onset of an experimental spoken word, and it stayed on screen for 2500 ms. After that, the drift correction resumed presentation until the next trial (See Figure 1 for sequence of trial events). The reason for including the preview period is to allow for non-linguistic visual searches such as location identification involved in the experimental task, so that visual-cognitive processes are not included to bias eye-tracking recordings (Apfelbaum et al., 2021). We set preview as 500 ms, a duration longer than that of which planning and executing an eye movement typically takes (200ms) as SI presumably involves more complex cognitive processes compared to word recognition tasks (as in Huettig & McQueen, 2007 and McQueen & Viebahn, 2007). The printed words were presented in 28-point SimSun font. They were displayed at the center of four fixed cells of a 7 × 7 grid (cells 17, 19, 31, 33 counting from left to right and top to bottom). The grid was introduced for design purpose and was not visible to participants (Huettig & McQueen, 2011). The positions of distractors and critical words were pseudo-randomized across trials. Critical words appeared on each of the four cells equally frequently.

Participants were instructed to listen to the spoken texts and perform speech production tasks, while keeping their eyes fixated on screen. They were told that they were free to view anything on the screen but should never move eyes off the screen throughout the experiment sessions. The speech production tasks consisted of shadowing and SI. Take the trial shown in Figure 1 as an example: In the SI task, participants were expected to produce the Chinese translation equivalent of the spoken English word (*buildings*—建筑 *[Jian Zhu]*) as they interpret the sentence, while being visually presented the competitor word (建议 *[Jian Yi]*) and three distractor words. In the shadowing task, they were asked to repeat the English sentence verbatim, irrespective of the visually presented Chinese words. The experiment started with a practice session in which participants listened to a 120-word long spoken text containing six trials. During this session, they were instructed to comprehend the speech while familiarizing themselves with the co-presence of visual and auditory information. Following the practice session, participants completed six speech production sessions, performing SI on three texts and performing shadowing on the other three. The order in which they performed the two tasks on the six texts was counterbalanced using a Latin square design. Immediately after each session, they completed a cued recall test. The entire experiment lasted for 60–70 minutes.

Data analysis

Eye-tracking data

Fixations were extracted using the Tobii Pro Lab software and prepared for analysis using the *eyetrackingR* package (Dink & Ferguson, 2015) and a collection of
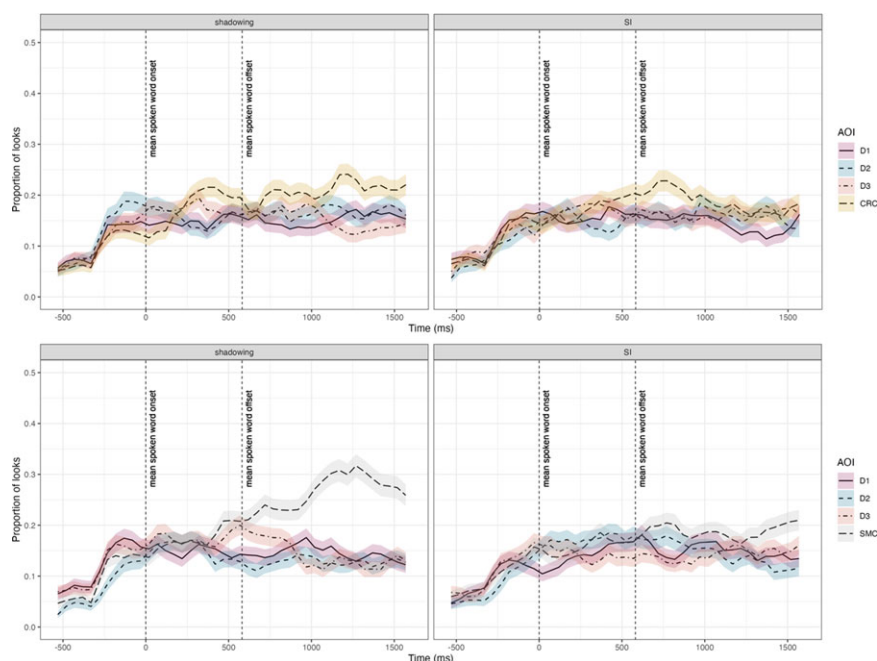
Figure 2. Time-course graph showing fixation proportions to four AOIs in character repetition condition (top) and semantic condition (bottom).
Notes: AOIs in character repetition condition consist of distractor 1 (D1), distractor 2 (D2), distractor 3 (D3) and character repetition competitor (CRC), and those in semantic condition consist of D1, D2, D3 and semantic competitor (SMC). The *x*-axis shows time in milliseconds from 500 ms before spoken word onset. Transparent thick lines represent ±1 standard error.

packages within the *tidyverse* (Wickham et al., 2019) in R. Each cell containing a written word was made an area of interest (AOI), and fixations were coded as fallen upon the competitor (i.e., character repetition competitor, CRC; semantic competitor, SMC), distractor 1 (D1), distractor 2 (D2), or distractor 3 (D3). As the average duration of spoken words is 581 ms, we included the time period of the words being spoken and 1000 ms succeeding word offset into analysis, amounting to a total of 1600 ms as critical time window. We made this decision to ensure that both early automatic processes and those involved in preparation for production (if existing) can be captured, since there is often a short time lag between input heard and output produced. Trials with over 25% trackloss were removed, resulting in 13.89% of data loss. For visualization purpose, the time period was divided into 50 ms time bins. For each time bin, fixation proportions were calculated for each AOI as the number of fixations relative to the total number of fixations (see Figure 2 for the distribution of fixation proportions of the four AOIs for all conditions).

Statistical analysis consisted of two major sections. First, time-bin analysis was conducted to assess fixation patterns within each time bin and to enable post hoc pairwise comparison and examine how different AOIs competed with each other in terms of attracting fixations as time unfolded. It should be noted that time bins lasting for 200 ms were included in analysis, which is consistent with previous visual

world studies (Huettig & McQueen, 2011; Prasad & Mishra, 2021), despite that the time bins were divided into 50 ms long for visualization. Fixation proportions to the three distractors were averaged into a single value, and all the proportion data were empirical logit transformed to accommodate for the bounded and categorical nature of data (fixate or not, coded as 0 or 1) before entering analysis (Barr, 2008). Separate analyses were conducted for the two conditions using linear mixed-effects models with the *lme4* package (Bates et al., 2015). For each time bin, the statistical model included fixed effects of AOI (competitor versus averaged distractors) and task (SI versus shadowing), all of which were sum-coded, and their interaction. By-subject and by-item random intercepts were also added (Barr, 2008). Statistical significance (*p*-values) was estimated from the *t* distribution.

Second, we ran growth curve analysis (GCA) with the *lme4* package, considering time as a continuous variable (Mirman, 2014). Although predictions were made about the speed of co-activations in different experimental conditions, we have no specific expectations regarding the shapes of gaze fixations unfolding across time. Therefore, we took a statistical approach to GCA modeling to improve model fit (Mirman, 2014). The non-linear changes in fixation proportions were modeled with fourth-order (quartic) orthogonal polynomial. The analysis tested for effects of condition and task, as well as an interaction between condition and task. We did not include fixations directed to distractors because our intention of using GCA is to verify the relative speed and magnitude of co-activations triggered by form and meaning. Fixation proportions to competitors were averaged over each 50 ms time bin and then went through empirical logit transformation. A linear mixed-effects model was constructed with trials included in regression weighted using the approach described by Mirman (2014, p. 111). The model consisted of fixed effects of task (SI vs. shadowing), condition (character repetition vs. semantic), and the interaction of the two on all the four time terms. Task and condition were both deviation coded. The transformed proportion data were taken as dependent variable. The model also included random effects of participants and trials on intercept and all the time terms. Random slopes were not included due to failure to converge in model. Improvements of model fit were determined via model comparison using –2 times the change in log-likelihood. Parameter-specific *p*-values were assessed using the normal approximation (Mirman, 2014).

Cued recall performance

Assessment method for performance on cued recall test was adopted from Christoffels and de Groot (2004). For each sentence with fragment, a score between 0 and 3 was given (0: no recall or error; 1: half finished; 2: similar in meaning; 3 completely correct recall), depending on the extent to which the recalled information was semantically close to the original content. Consistent with Christoffels and de Groot (2004), we focused on recall of meaning rather than verbatim recall. After receiving a brief training on the rating system and criteria, two independent raters assessed participants' performance. The inter-rater reliability was acceptable, $\rho = 0.81$ (> 0.70). A paired *t*-test was performed on averaged ratings for the two tasks. Participants scored

significantly higher on cued recall after SI (*Mean* = 55.70, *SD* = 9.58) than shadowing (*Mean* = 48.78, *SD* = 11.57), $t(26)$ = −2.82, p = .009. The effect size, as measured by Cohen's *d*, was 0.65 with a 95% confidence interval [0.10, 1.20].

Results

Time-bin analysis

Table 2 shows a summary of statistical results for the character repetition condition. Differences in the looks to AOIs initially appeared around 400–600 ms, with both groups showing more looks to the CRC than to averaged distractors (β = 0.04, SE = 0.02, t = 2.45, p = .014). The effects of AOI were present toward the end of the time window (600–800 ms: β = 0.03, SE = 0.02, t = 2.09, p = .037; 800–1000 ms: β = 0.03, SE = 0.02, t = 1.98, p = .048; 1000–1200 ms: β = 0.04, SE = 0.02, t = 2.85, p = .004; 1200–1400 ms: β = 0.05, SE = 0.02, t = 3.18, p = .002; 1400–1600 ms: β = 0.05, SE = 0.02, t = 2.74, p = .006). No effects of task or interaction effects were found throughout the time bins analyzed.

Table 3 shows a summary of statistical results for the semantic condition. Differences in AOI began to show in the 400–600 ms time bin and lasted towards the end of the time window, with participants looking more at the SMC than averaged distractors (400–600 ms: β = 0.05, SE = 0.02, t = 2.92, p = .004; 600–800 ms: β = 0.07, SE = 0.02, t = 4.19, p <.001; 800–1000 ms: β = 0.10, SE = 0.02, t = 5.83, p <.001; 1000–1200 ms: β = 0.15, SE = 0.02, t = 8.67, p <.001; 1200–1400 ms: β = 0.14, SE = 0.02, t = 8.32, p <.001; 1400–1600 ms: β = 0.13, SE = 0.02, t = 7.19, p <.001). The AOI by task interaction effect was significant in the 400–600 ms time bin (β = −0.05, SE = 0.02, t = −2.08, p = .038) as well as the time bins from 800 ms to 1600 ms (800–1000 ms: β = −0.06, SE = 0.03, t = −2.26, p = .024; 1000–1200 ms: β = −0.12, SE = 0.03, t = −4.86, p <.001; 1200–1400 ms: β = −0.11, SE = 0.03, t = −4.28, p <.001; 1400–1600 ms: β = −0.06, SE = 0.03, t = −2.22, p = .027).

Planned comparisons indicated that there were significantly more looks to the SMC than distractors in the shadowing task in 400–600 ms (β = −0.05, SE = 0.02, t = −2.92, p = .019) and in 800–1000 ms (β = −0.10, SE = 0.02, t = −5.84, p < .001). The interaction effects found during 1000–1400 ms were due to more fixations to the SMC than distractors in shadowing (1000–1200 ms: β = −0.15, SE = 0.02, t = −8.67, p < .001; 1200–1400 ms: β = −0.14, SE = 0.02, t = −8.32, p < .001) and more fixations to the SMC in shadowing than in SI (1000–1200 ms: β = 0.09, SE = 0.02, t = 5.14, p < .001; 1200–1400 ms: β = 0.08, SE = 0.02, t = 4.52, p < .001). As for the last time bin, the interaction effect was because of significant competitor-distractor differences in both shadowing (β = −0.13, SE = 0.02, t = −7.20, p < .001) and SI (β = −0.07, SE = 0.02, t = −3.61, p = .002).

Growth curve analysis

To test for whether co-activation driven by word form differs from that driven by meaning, we modeled the proportion data of both competitors using GCA. The

Table 2. Results of LME time-bin analysis for character repetition condition

	0 ms – 200 ms		200 ms – 400 ms		400 ms – 600 ms		600 ms – 800 ms	
	Est.	*t*	*Est.*	*t*	*Est.*	*t*	*Est.*	*t*
Intercept	−0.15 (0.01)	−13.31***	−0.16 (0.01)	−14.45***	−0.17 (0.01)	−14.85***	−0.16 (0.01)	−14.29***
Task type	−0.00 (0.02)	−0.27	0.01 (0.02)	0.49	0.00 (0.02)	0.17	−0.01 (0.02)	−0.36
AOI	−0.02 (0.02)	−1.12	0.03 (0.02)	1.80	0.04 (0.02)	2.45*	0.03 (0.02)	2.09*
Task type × AOI	0.02 (0.02)	0.74	−0.03 (0.02)	0.17	−0.01 (0.02)	−0.49	0.02 (0.02)	1.01
	800 ms – 1000 ms		1000 ms – 1200 ms		1200 ms – 1400 ms		1400 ms – 1600 ms	
	Est.	*t*	*Est.*	*t*	*Est.*	*t*	*Est.*	*t*
Intercept	−0.16 (0.01)	−14.59***	−0.17 (0.01)	−15.03***	−0.17 (0.01)	−14.52***	−0.17 (0.01)	−14.05***
Task type	0.00 (0.02)	0.02	0.01 (0.02)	0.36	0.01 (0.02)	0.32	0.00 (0.02)	0.23
AOI	0.03 (0.02)	1.98*	0.04 (0.02)	2.85**	0.05 (0.02)	3.18**	0.05 (0.02)	2.74**
Task type × AOI	−0.00 (0.02)	−0.05	−0.02 (0.02)	−1.03	−0.02 (0.03)	−0.92	−0.02 (0.02)	−0.64

*Notes*: AOI = area of interest; Est. = model estimate (standard error in brackets); * = $p < .10$, ** = $p < .05$, *** = $p < .001$.

Table 3. Results of LME time-bin analysis for semantic condition

	0 ms – 200 ms		200 ms – 400 ms		400 ms – 600 ms		600 ms – 800 ms	
	Est.	*t*	*Est.*	*t*	*Est.*	*t*	*Est.*	*t*
Intercept	−0.16 (0.01)	−13.32***	−0.16 (0.01)	−14.67***	−0.17 (0.01)	−15.09***	−0.17 (0.01)	−14.79***
Task type	0.00 (0.02)	0.00	0.00 (0.02)	0.17	0.01 (0.02)	0.74	0.01 (0.02)	0.45
AOI	0.01 (0.02)	0.49	0.01 (0.02)	0.89	0.05 (0.02)	2.92**	0.07 (0.02)	4.19***
Task type × AOI	−0.00 (0.02)	−0.16	−0.01 (0.02)	−0.48	−0.05 (0.02)	−2.08*	−0.03 (0.02)	−1.27
	800 ms – 1000 ms		1000 ms – 1200 ms		1200 ms – 1400 ms		1400 ms – 1600 ms	
	Est.	*t*	*Est.*	*t*	*Est.*	*t*	*Est.*	*t*
Intercept	−0.18 (0.01)	−15.08***	−0.19 (0.01)	−16.15***	−0.19 (0.01)	−15.80***	−0.19 (0.01)	−14.37***
Task type	0.01 (0.02)	0.80	0.03 (0.02)	1.72	0.03 (0.02)	1.52	0.02 (0.02)	0.79
AOI	0.10 (0.02)	5.83***	0.15 (0.02)	8.67***	0.14 (0.02)	8.32***	0.13 (0.02)	7.19***
Task type × AOI	−0.06 (0.03)	−2.26*	−0.12 (0.03)	−4.86***	−0.11 (0.03)	−4.28***	−0.06 (0.03)	−2.22*

Notes: AOI = area of interest; Est. = model estimate (standard error in brackets); * = $p < .10$, ** = $p < .05$, *** = $p < .001$.
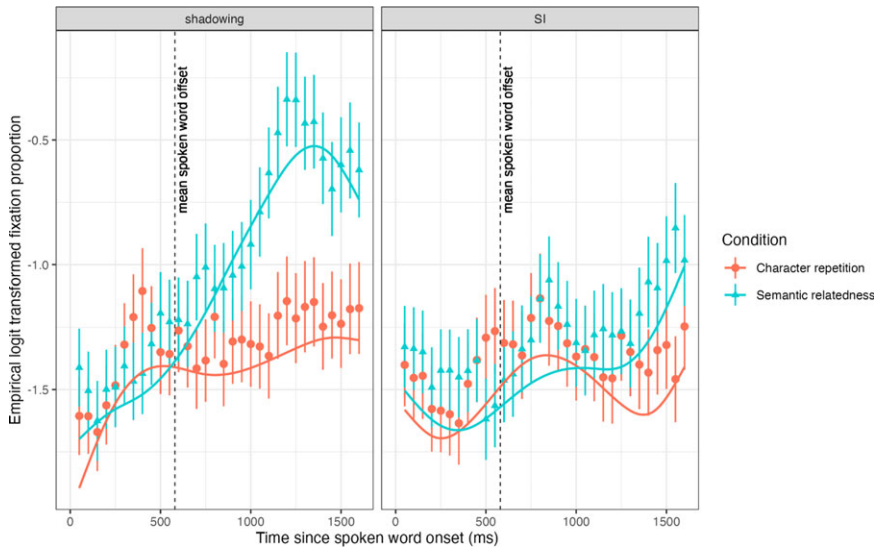
Figure 3. Growth curve model fits (lines) and observed empirical logit transformed fixation proportion data (dots) for the effect of condition and task.
Notes: Error bars represent ±1 standard error.

model revealed a significant effect of task (β = –0.23, SE = 0.03, t = –7.21, p = .000) on the intercept term, suggesting there were more looks to competitors in shadowing than in SI overall. The effect of task was also significant on the linear term (β = –0.83, SE = 0.18, t = –4.63, p = .000) and on the quartic term (β = 0.69, SE = 0.18, t = 3.89, p = .000). An inspection of Figure 3 shows that the effect on the linear term was because the looks to competitors in shadowing kept climbing overall, while those in SI increased slightly. The effect on the quartic term was due to the fact that the tendency of proportion data in SI showed a quartic curve overall, which was absent in shadowing. The effects of task indicate that shadowing yielded significantly more co-activations, and the effects occurred slowly and steadily and in an increasing way, particularly in semantic condition. In SI, the effects were more sensitive to the spoken word offset, with increase of looks observed shortly after spoken word offset and decreased fixations after that. The condition by task interaction effect was significant on the intercept term (β = –0.26, SE = 0.06, t = –4.24, p = .000) and on the linear term (β = –1.05, SE = 0.35, t = –2.97, p = .003). This is because looks to SMC in shadowing increased greatly after spoken word offset and maintained growing compared to CRC, whereas no such difference between conditions was observed for SI. The interaction effect was also significant on the cubic term (β = –0.17, SE = 0.35, t = 2.41, p = .016), which was largely because of the different shape and temporal unfolding of the peaked curves. As shown in Figure 3, looks to CRC in shadowing reached a peak approximately around 400 ms (earlier than word offset), and a huger peak was observed around 1300 ms for the semantic condition. In contrast, SI did not yield a peak for the semantic condition, while looks to CRC showed a peak shortly after spoken word offset, later than the peak observed in shadowing.

There were two inverted peaks in the curve of looks to CRC in SI, which was absent in shadowing.

Discussion

The present study aimed to explore cross-language co-activation during SI. We examined whether the conscious use of both languages and the frequent switching between them influenced co-activation patterns as compared to shadowing. We also established a GCA model to examine the time course of cross-language co-activation as modulated by the relatedness in word forms and meanings. Our results showed that, when hearing the critical English word, interpreters were more likely to look at competitors sharing characteristics with their Chinese translation equivalent than unrelated distractors regardless of tasks or conditions. An increase in the proportion of fixations to character repetition competitors was observed in the time bin pertaining to the spoken word offset (i.e., 400–600 ms), an effect that was consistent across the two tasks, while the competitor effect found in the semantic condition occurred earlier and in a stronger magnitude in shadowing than in SI. We discuss the results by comparing tasks and conditions with reference to our first prediction (Section 4.1) and second prediction (Section 4.2).

The effect of tasks

In the present study, we observed parallel lexical activation in a task that features intensive code-switching in both comprehension and production. Participants fixated considerably more on both character repetition and semantic competitors approximately upon spoken word offset, suggesting spontaneous activation of orthographic and semantic features of the Chinese translation equivalent when receiving phonological information of the English word. Previous studies made the case that both languages were constantly active even when one language was not in use (Thierry & Wu, 2007; Lagrou et al., 2011). We add to the strong account of language non-selective lexical access in bilinguals by showcasing that parallel activation plays a role as critical in cognitively taxing tasks like SI as in other bilingual tasks involving only one language.

In our prediction for task comparison, the magnitude of co-activation should be strengthened in SI by interpreters' overt intention to translate and switch between languages, as compared to shadowing. Our results did not support this prediction. In the character repetition condition, the competitor-distractor difference did not yield an interaction effect, which means that the magnitude of cross-language co-activation driven by orthographic and phonological similarity was comparable across tasks. Although participants were consciously directed to target-language articulation in SI, the extent of automatic activation occurred was not different from that observed in shadowing. The absence of task difference suggests that the word form of the translation equivalent can be immediately activated upon receiving speech signal regardless of the language used for production. In other words, the top-down intention to switch between languages does not modulate early-stage, bottom-up lexical access.

Our results are contradictory to those of Amos et al. (2022) showing no evidence for word-form activation during SI with English-French interpreters. The authors, who intended to capture prediction effects of phonological information in their design though, observed a lack of competitor-distractor difference not only in the prediction time window but even after the spoken word onset. A possible explanation for the discrepancy was that the experimental sentences used in their study were highly contextually constrained, a design intended to trigger anticipation of upcoming words. In their case, the linguistic context confined bottom-up activation of word form, as participants were drawn to semantic processing considerably. In our study, we controlled the predictability of critical words which were neutral in the given sentential contexts, leading to unbiased attention on form and meaning. The contrast between our findings and those of Amos and colleagues lends support to one of BIA+'s hypotheses that the linguistic context modulates non-selectivity (Dijkstra & van Heuven, 2002). Moreover, our results also suggest that a challenging task such as the SI does not necessarily prevent bottom-up activation of word form.

For the semantic condition, we found significantly larger competitor-distractor differences in shadowing than in SI around 1000–1600 ms. The result seems to suggest that shadowing involved a more profound level of semantic processing, which is at odds with previous evidence for the nature of semantic processes involved in self-paced word repetition (Gustafson et al., 2013) and sentence shadowing (Christoffels & de Groot, 2004). Meanwhile, the recall performance in SI was better than in shadowing, suggesting that the input was processed deeper in SI. We propose two explanations for these findings. First, given that there is typically a short time lag between the input and output in both tasks, it is reasonable to believe that articulation of the translation equivalents of the words heard, rather than auditory perception alone, also contributes to increased semantic processing. We reasoned that relatively lower visual attention on semantic competitors in SI compared to shadowing was due to the fact that the delay between comprehension and production in shadowing is often shorter than that in SI (Timarová et al., 2011). In fact, the production latencies calculated in the performances of the two tasks in our study were an average of 1.10 seconds for shadowing ($SD = 0.35$) and 3.14 seconds for SI ($SD = 0.68$). In this case, the strong semantic activation found for the shadowing task could likely be driven by concurrent production. Interestingly, this effect was absent in the character repetition condition, suggesting that the co-activation of word form in the other language is unlikely to be involved in single-language production such as shadowing. Our finding is in line with Strijkers et al.'s (2011) point of view that top-down intentions to speak proactively facilitate the perception of the meaning of a word.

Second, less visual attention on semantic competitors during SI may not necessarily suggest a lower level of semantic engagement. It could be due to heightened cognitive load exerted by the task itself, as participants had to listen, reformulate, and produce target renditions all at the same time. They must quickly disengage from comprehended information and prepare for language-switching and production, taking into consideration lexico-semantic equivalence and sentence restructuring in the meantime. Therefore, participants' visual attention could be largely constrained as a result of constantly engaging in overlapped language tasks

under extreme time pressure. Nonetheless, they still achieved good performances at the cued recall test. As suggested by Christoffels and de Groot (2004), the intention of translation could serve as a cue to recall. It is likely that participants engaged in the semantic meanings of the spoken texts more deeply because of the intention to translate, while the semantic engagement was less profound during shadowing.

The time course of cross-language co-activation

We found an early effect of AOI (appearing since 400 ms) in the character repetition condition irrespective of tasks. Participants quickly directed their attention towards the orthographic neighbor of the translation equivalent compared to distractors, suggesting activation of the form of the word in the Chinese language through implicit translation which further spread to activation of other lexical candidates. Our results expand earlier findings with cross-script language pairs that have shown early attention bias towards competitors sharing orthographic similarity with translation equivalents (Mishra & Singh, 2014; Prasad & Mishra, 2021). However, word-form co-activations of the unspoken language have only been demonstrated in participants performing L2 spoken word recognition tasks in sentences. Our study is the first to show the same effect when bilinguals were engaged in a continuous speech involving both comprehension and production in a synchronized fashion with short time lags.

The early character repetition competitor effect is similar to that observed by Thierry and Wu (2007) in the classic N400 time window (i.e., 400 ms after the onset of stimuli) in non-interpreter Chinese-English bilinguals. The consistency in the temporal onset of the effects indicates the same mechanism underlying cross-language interactions in L2 comprehension supporting tasks involving concurrent and (explicit) cross-language comprehension and production. However, the N400 character repetition effect was short and transient in Thierry and Wu (2007), while the effect was more durable in our study, existing until 1600 ms. In addition to evidence supporting non-selective access in recognition tasks, parallel activation of unused language via cues of word form (i.e., cognate words) has been found in L2 production tasks (Starreveld et al., 2014) which might suggest that the top-down modulation involved in preparation for production influenced the selection of language, but this effect was equally visible in shadowing, a task involving no switch of languages between comprehension and production. The absence of task difference suggests that the activation of translation equivalent is independent of whether cross-language switching is involved in the task or not.

Similar to the case of character repetition, the competitor effect in the semantic condition was first observed in the 400–600 ms time bin and lasted until the end of time window. In cross-task comparison, however, the effect was only present during SI in the 600–800 ms and 1400–1600 ms bins, significantly smaller than those observed in shadowing. As previously explained, the task difference in the semantic competitor effect could be due to 1) the semantic mediation involved in the production of shadowing, a task with relatively shorter production latency, and 2) increased cognitive load in SI. More importantly, the earlier appearance of the word-form competitor effect relative to the semantic competitor effect is consistent with our prediction that parallel activation of the target language serves as a

dominant mechanism in SI processing. The finding is also in line with the hypothesis of the BIA+ model that word forms are processed preconsciously at an early stage followed by top-down, meaning-driven processing. However, the fact that semantic activation took place in two distinct time bins—first following spoken word offset and later at a stage possibly linked to preparation for production— seems to suggest less synchronized comprehension and production, in other words, a serial processing mode.

The different patterns of co-activations across the two languages driven by word-form repetition and semantic association have significant implications for the parallel/serial processing account of SI. Our findings revealed a pattern of early, bottom-up word-form activation and a stepwise semantic activation during SI. Consistent with our prediction, our findings support the parallel view, that is, direct mapping of translation equivalents can be achieved at least at lexical level, as we found earlier co-activation of target-language word-form information than that of semantic meaning. However, this does not mean that spontaneous, bottom-up activation of the target language could facilitate conscious, deliberate processing of translation equivalents. First, no stronger character repetition effect was observed for SI compared to shadowing, as discussed in section 4.1. Second, SI involved serial processing of the semantic information (i.e., early activation and re-activation at a later time), which suggests that early parallel activation of word form and meaning did not feedforward to the production stage. Instead, the semantic meaning was re-processed separately when preparing for production. Meanwhile, the by-the-quartic-term effect of task in the GCA model indicates that the competitors in both conditions attracted visual attention at two separate time points when SI was being performed, while such quartic pattern was absent in shadowing. The evidence combined points to that the small latencies between input and output have a substantial effect on the time course of cross-language interactions. The simultaneity of perceiving and speaking, which is the characteristic of SI, is less precisely described in previous conceptualizations of interpreting (e.g., Christoffels & de Groot, 2009; Dong & Li, 2019; Giles, 2009). In fact, comprehension and production in the task were seldom synchronized. As a result, the cross-language effects taking place during SI could involve more complicated mechanisms than implicated by parallel and serial processing models.

Conclusion

We investigated cross-language co-activation in SI and shadowing. The results showed that both the meaning of a word and its translation equivalent form were activated in an early stage of perception, with word-form activations occurring consistently from 400 ms following spoken word onset until the end of time window while semantic activations took place in two separate time bins. Our findings suggest that cross-language co-activation during concurrent comprehension and production was instant and spontaneous as soon as acoustic information of spoken word was received, whereas the time lag between comprehension and production modulated the time course of co-activation. A dichotomous view of SI processing as either parallel or serial seems oversimplified as an account for the

complexities involved in the task. Investigation into the relationship between cross-language co-activation and the time lag in SI performance could potentially contribute to this discussion. In addition, both SI and shadowing yielded the same pattern of word-form activation, which implies that the pattern was unaffected by whether the same language was consistently used for comprehension and production. This conclusion is not yet generally applicable to all bilinguals since we only included professional interpreters as participants in the experiment. There might be an effect of expertise or training in SI that transfers to the shadowing task. Comparisons between experienced and novice interpreters, as well as with non-interpreter bilinguals, are needed for future studies. It is also unknown whether or not co-activations of form and meaning mutually facilitate each other, as suggested in the cascaded view of language-mediated visual attention (Huettig & McQueen, 2007). An experimental design that examines both character repetition competitor and semantic competitor in the same visual display could better illustrate how co-activations of form and meaning compete for visual attention as time unfolds.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S014271642510009X

Replication package. The stimulus materials, data, and analysis scripts can be found via the Open Science Framework at https://doi.org/10.17605/OSF.IO/RCEDW.

Competing interests. We have no known conflict of interest to disclose.

Note

1 In interpreting, "target language" refers to the language that one translates into. The term, however, is often used to indicate the language that is being used in the context of bilingual processing. To distinguish, we use "source language" and "target language" to suggest the languages that are translated from and into respectively, while "language in use" and "unused language" are used to suggest the explicitly used and unused languages in regular bilingual tasks.

References

Amos, R. M., Seeber, K. G., & Pickering, M. J. (2022). Prediction during simultaneous interpreting: Evidence from the visual-world paradigm. *Cognition*, **220**, 104987. https://doi.org/10.1016/j.cognition.2021.104987

Andras, F., Rivera, M., Bajo, T., Dussias, P. E., & Paolieri, D. (2022). Cognate facilitation effect during auditory comprehension of a second language: A visual world eye-tracking study. *International Journal of Bilingualism*, **26**(4), 405–425. https://doi.org/10.1177/13670069211033359

Apfelbaum, K. S., Klein-Packard, J., & McMurray, B. (2021). The pictures who shall not be named: Empirical support for benefits of preview in the Visual World Paradigm. *Journal of Memory and Language*, **121**, 104279. https://doi.org/10.1016/j.jml.2021.104279

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, **59**(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, **22**(5), 633–660. https://doi.org/10.1080/01690960601000746

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, **46**(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **35**(4), 1029–1040. https://doi.org/10.1037/a0015901

Chen, P., Bobb, S. C., Hoshino, N., & Marian, V. (2017). Neural signatures of language co-activation and control in bilingual spoken word comprehension. *Brain Research*, **1665**, 50–64. https://doi.org/10.1016/j.brainres.2017.03.023

Christoffels, I. K., & de Groot, A. M. B. (2004). Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition*, **7**(3), 227–240. https://doi.org/10.1017/S1366728904001609

Christoffels, I. K., & de Groot, A. M. B. (2009). Simultaneous interpreting: A cognitive perspective. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 454–479). New York: Oxford University Press.

Covington, M. A. (2009). Idea density: A potentially informative characteristic of retrieved documents. *IEEE Southeastcon*, **2009**, 201–203. https://doi.org/10.1109/SECON.2009.5174076

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, **41**(4), 496–518. https://doi.org/10.1006/jmla.1999.2654

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, **42**(4), 445–464. https://doi.org/10.1006/jmla.1999.2697

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, **5**(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dink, J., & Ferguson, B. (2015). *eyetrackingR: An R Library for Eye-tracking Data Analysis*. http://www.eyetrackingr.com

Dong, Y., & Li, P. (2019). Attentional control in interpreting: A model of language control and processing control. *Bilingualism: Language and Cognition*, **23**(4), 716–728. https://doi.org/10.1017/S1366728919000786

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **33**(4), 663–679. https://doi.org/10.1037/0278-7393.33.4.663

FitzPatrick, I., & Indefrey, P. (2014). Head start for target language in bilingual listening. *Brain Research*, **1542**, 111–130. https://doi.org/10.1016/j.brainres.2013.10.014

Gile, D. (2009). *Basic concepts and models for interpreter and translator training* (Rev. ed.). Amsterdam: John Benjamins.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, **36**(2), 193–202. https://doi.org/10.3758/BF03195564

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, **25**(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Grosjean, F. (1998). Transfer and language mode. *Bilingualism: Language and Cognition*, **1**(3), 175–176. https://doi.org/10.1017/S1366728998000285

Gustafson, E., Engstler, C., & Goldrick, M. (2013). Phonetic processing of non-native speech in semantic vs non-semantic tasks. *The Journal of the Acoustical Society of America*, **134**(6), EL506–EL512. https://doi.org/10.1121/1.4826914

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, **57**(4), 460–482. https://doi.org/10.1016/j.jml.2007.02.001

Huettig, F., & McQueen, J. M. (2011). The nature of the visual environment induces implicit biases during language-mediated visual search. *Memory & Cognition*, **39**(6), 1068–1084. https://doi.org/10.3758/s13421-011-0086-z

Jiang, N. (2021). Examining L1 influence in L2 word recognition: A case for case. *Journal of Second Language Studies*, **4**(1), 1–18. https://doi.org/10.1075/jsls.19039.jia

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, **31**(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, **44**(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **37**(4), 952–965. https://doi.org/10.1037/a0023217

Liu, Y., Hintz, F., Liang, J., & Huettig, F. (2022). Prediction in challenging situations: Most bilinguals can predict upcoming semantically-related words in their L1 source language when interpreting. *Bilingualism: Language and Cognition*, **25**(5), 801–815. https://doi.org/10.1017/S1366728922000232

Lozano-Argüelles, C., & Sagarra, N. (2021). Interpreting experience enhances the use of lexical stress and syllabic structure to predict L2 word endings. *Applied Psycholinguistics*, **42**(5), 1135–1157. https://doi.org/10.1017/S0142716421000217

Lozano-Argüelles, C., Sagarra, N., & Casillas, J. V. (2020). Slowly but surely: Interpreting facilitates L2 morphological anticipation based on suprasegmental and segmental information. *Bilingualism: Language and Cognition*, **23**(4), 752–762. https://doi.org/10.1017/S1366728919000634

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, **88**, 22–60. https://doi.org/10.1016/j.cogpsych.2016.06.002

Macizo, P., Bajo, T., & Cruz Martín, M. (2010). Inhibitory processes in bilingual language comprehension: Evidence from Spanish-English interlexical homographs. *Journal of Memory and Language*, **63**(2), 232–244. https://doi.org/10.1016/j.jml.2010.04.002

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, **244**(5417), 522–523. https://doi.org/10.1038/244522a0

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, **60**(5), 661–671. https://doi.org/10.1080/17470210601183890

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Taylor & Francis.

Mishra, R. K., & Singh, N. (2014). Language non-selective activation of orthography during spoken word processing in Hindi–English sequential bilinguals: An eye tracking visual world study. *Reading and Writing*, **27**(1), 129–151. https://doi.org/10.1007/s11145-013-9436-5

Paradis, M. (1994). Towards a neurolinguistic theory of simultaneous translation: The framework. *International Journal of Psycholinguistics*, **10**, 319–335.

Prasad, S., & Mishra, R. K. (2021). Concurrent verbal working memory load constrains cross-linguistic translation activation: A visual world eye-tracking study on Hindi-English bilinguals. *Bilingualism: Language and Cognition*, **24**(2), 241–270. https://doi.org/10.1017/S1366728920000024

Seleskovitch, D. (1976). Interpretation: A psychological approach to translating. In R. W. Brislin (Ed.), *Translation: Applications and research* (pp. 92–116). New Yorker: Gardiner.

Shook, A., & Marian, V. (2019). Covert co-activation of bilinguals' non-target language: Phonological competition from translations. *Linguistic Approaches to Bilingualism*, **9**(2), 228–252. https://doi.org/10.1075/lab.17022.sho

Starreveld, P. A., De Groot, A. M. B., Rossmark, B. M. M., & Van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism: Language and Cognition*, **17**(2), 258–276. https://doi.org/10.1017/S1366728913000308

Strijkers, K., Holcomb, P. J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal of Memory and Language*, **65**(4), 345–362. https://doi.org/10.1016/j.jml.2011.06.002

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, **104**(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Timarová, Š., Dragsted, B., & Hansen, I. G. (2011). Time lag in translation and interpreting. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 121–146). John Benjamins Publishing Company.

Türker, E. (2018). The influence of L1 frequency in instructed second language learning of L2 idioms. *Journal of Second Language Studies*, **1**(2), 333–356. https://doi.org/10.1075/jsls.17007.tur

Valente, D., Ferré, P., Soares, A., Rato, A., & Comesaña, M. (2018). Does phonological overlap of cognate words modulate cognate acquisition and processing in developing and skilled readers? *Language Acquisition*, **25**(4), 438–453. https://doi.org/10.1080/10489223.2017.1395029

Van Hell, J. G., & de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, **128**(3), 431–451. https://doi.org/10.1016/j.actpsy.2008.03.010

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, **9**(4), 780–789. https://doi.org/10.3758/BF03196335

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, **50**(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, **4**(43), 1686. https://doi.org/10.21105/joss.01686

Wu, Y. J., & Thierry, G. (2010). Chinese–English bilinguals reading English hear Chinese. *The Journal of Neuroscience*, **30**(22), 7646–7651. https://doi.org/10.1523/JNEUROSCI.1602-10.2010

Wu, Y. J., & Thierry, G. (2012a). How reading in a second language protects your heart. *The Journal of Neuroscience*, **32**(19), 6485–6489. https://doi.org/10.1523/JNEUROSCI.6119-11.2012

Wu, Y. J., & Thierry, G. (2012b). Unconscious translation during incidental foreign language processing. *NeuroImage*, **59**(4), 3468–3473. https://doi.org/10.1016/j.neuroimage.2011.11.049

Wu, Y. J., & Thierry, G. (2017). Brain potentials predict language selection before speech onset in bilinguals. *Brain and Language*, **171**, 23–30. https://doi.org/10.1016/j.bandl.2017.04.002