A Network Science Approach to Bilingual Code-switching

Qihui Xu
Department of Psychology
Graduate Center, CUNY
qxu@gradcenter.cuny.edu

Magdalena Markowska
Department of Linguistics
Institute for Advanced Computational Science
Stony Brook University
magdalena.markowska@stonybrook.edu

Martin Chodorow
Department of Psychology
Hunter College, CUNY
mchodoro@hunter.cuny.edu

Ping Li
Department of Chinese and Bilingual Studies
The Hong Kong Polytechnic University
ping2.li@polyu.edu.hk

Abstract

Previous research has shown that the structure of the semantic network can influence language production, such that a word with low clustering coefficient (C) is more easily retrieved than a word with high C. In this study, we used a network science approach to examine whether the network structure accounts for why bilinguals code-switch. We established semantic networks for words in each language, then measured the C for each code-switched word and its translated equivalent. The results showed that words where language is switched have lower C than their translated equivalents in the other language, suggesting that the structures of the lexicons in the two languages play an important role in bilingual code-switching speech.

1 Background

Code-switching is defined as "the alternation between two (or more) languages within a single discourse, sentence, or constituent" (Poplack, 2000). The phenomenon has been widely observed in bilinguals with a variety of different language pairs (Poplack, 2000; Santorini and Mahootian, 1995; Barman et al., 2014). Understanding the mechanism behind bilingual code-switching speech is of particular importance to psycholinguistics and natural language processing. In psycholinguistics, it provides key clues to bilinguals' coordination between the representations of two or more languages from structural, psychological and social perspectives (Bullock and Toribio, 2009). In natural language processing, on the other hand, its mechanisms could provide a strong foundation for tasks, such as language identification (Lyu et al., 2006; Barman et al., 2014) and syntactic parsing (Broersma, 2009), among others. This study proposes and tests an explanation of why bilingual speakers switch their codes, and, in doing so, it introduces a new possible angle for code-switching research.

While code-switching could be analyzed at the sentence-level from the syntactic (Poplack, 2000; Santorini and Mahootian, 1995) and pragmatic perspectives (Beebe and Giles, 1984; Myers-Scotton, 1993), or at the phrase-level (Couto and Gullberg, 2019), our study focuses on a word-level account. Lexical properties of words, such as frequency, length, concreteness, and imageability (Gross and Kaushanskaya, 2015; Marian, 2009; Gollan et al., 2014; Gollan and Ferreira, 2009) may account for code-switching speech. Among these, word frequency has been characterized as a classical indicator of lexical accessibility that influences language choice in code-switching (Gollan and Ferreira, 2009; Gross and Kaushanskaya, 2015; Gollan et al., 2014). For example, Gross and Kaushanskaya (2015) found that words with a higher frequency of use for bilingual children were more likely to be produced than corresponding translations in the other language.

Strikingly, almost all previous studies of word-level code-switching have focused on a local rather than a global context. A local context treats each word as being independent from other words, whereas a global context emphasizes the interconnections and the interactions between words (Karuza et al., 2016). In addition to the attributes of a word itself, such as its frequency and concreteness, the global system of words has also been shown to affect many aspects of language processing (Hills et al., 2009; Sizemore et al., 2018; Steyvers and Tenenbaum, 2005; Chan and Vitevitch, 2009, 2010). In the domain of bilingual code-switching, however, little has been done to understand how a word, in a global interconnected lexical system, can affect the process of lexical selection. In this study, we address this question through a network science approach.

Proceedings of the Society for Computation in Linguistics (SCiL) 2021, pages 18-27.
Held on-line February 14-19, 2021

In the following sections, we review theoretical background of how a word and its associations with other words could potentially account for code-switching speech. First, we introduce the fundamental theory of what drives bilinguals to switch codes automatically, namely, the accessibility-driven account of (Kleinman and Gollan, 2016). Next, we review how the structure of the network of words moderates the accessibility of those words. Finally, a bridge between the semantic network structure and code-switching speech is proposed by introducing a particular structural measure (i.e. the clustering coefficient) which is used to examine the proposed account.

1.1 Accessibility-driven switching

Empirical evidence has shown that the process of code-switching can be effortless (Li, 1996; Gollan and Ferreira, 2009; Kleinman and Gollan, 2016). When bilinguals switch back and forth between languages in a free conversation, they can often do so as efficiently as communicating solely in one language, without any additional cognitive cost.

Other studies have shown that the switching of the code is driven by prior information (Li, 1996), which was has been referred to as the 'accessibility' of the code (Kleinman and Gollan, 2016; de Bruin et al., 2018; Gollan et al., 2014; Gross and Kaushanskaya, 2015). de Bruin et al. (2018) found that when doing voluntary code-switching, participants would choose the word that could be produced faster in a previous picture naming task than the corresponding word in the other language. As more accessible words are believed to be produced faster when being cued, the authors argued that language choice during code-switching is driven by lexical accessibility.

Previous evidence for the accessibility account comes mainly from word-by-word switching in experimental settings, which is ecologically dissimilar to the intrasentential code-switching found in a spontaneous speech. As a result, the present study examines code-switching speech recorded in natural conversations.

1.2 Words in network science

Network science has been increasingly applied in the research fields of linguistics and cognitive science. It aims to answer questions about how immensely complex pieces of information interact with each other and how they shape human cognition and behaviors (Karuza et al., 2016). A large number of studies have revealed the influence of network structure on language acquisition (Sizemore et al., 2018; Hills et al., 2009), representation (Steyvers and Tenenbaum, 2005) and production (Goldrick and Rapp, 2007; Chan and Vitevitch, 2010).

Semantic networks, an important domain in network science, provide the means of organizing and representing the meanings of words. In a typical semantic network, words are represented as nodes, and the semantic associations between them are represented as edges. In a weighted semantic network, the weight of an edge is determined by the strength of the shared semantic association between the two words. Figure 1 illustrates a weighted semantic network.

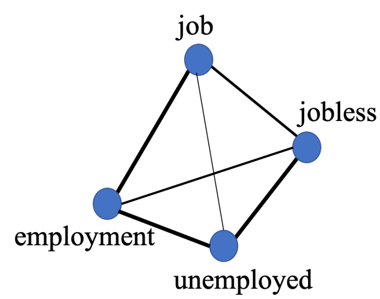

Figure 1: An illustration of a weighted semantic network of four semantically-related words. Each node represents a word, while the edge linking the nodes represents the semantic association between those words. The weight of the edge is determined by the semantic similarity between the two linked words. The thicker lines in the graph indicate higher semantic similarity between the two words.

Semantic networks have provided a mechanistic account for conducting research in various domains of language research, such as first language acquisition (Hills et al., 2009; Sizemore et al., 2018), lexical memory (Vitevitch et al., 2012), word recognition (Yates, 2013), to name a few. By comparing the properties of a semantic network and human behavior, researchers have also found that various properties of the network may be correlated with human language processing and representation (Vitevitch et al., 2012; Goldrick and Rapp, 2007; Storkel et al., 2006). One such network property that is examined in this study is the clustering coefficient (Watts and Strogatz, 1998), which is a commonly-used standard measurement for describing network structure that has been shown to influence the accessibility of the spoken words (Siew, 2019; Chan and Vitevitch, 2009, 2010; Vitevitch et al., 2011).

Clustering coefficients The clustering coefficient (C) represents the probability that two neighbors of a node are themselves neighbors (Watts and Strogatz, 1998). C is measured based on triangle subgraphs, which represent the interconnections between a node and any of its two neighbors. More specifically, C of a word in a weighted network is calculated by taking the sum of the geometric average of the subgraph edge weights devided by all the possible connections that the word could have with its neighboring words (Onnela et al., 2005):

$$C_u = \frac{2}{deg(u)(deg(u)-1)} \sum_{vx} (\hat{W}_{uv}\hat{W}_{ux}\hat{W}_{vx})^{\frac{1}{3}} \quad (1)$$

where $deg(u)$ represents the number of edges a given node is connected to. $\hat{W}_{uv}$, $\hat{W}_{ux}$, and $\hat{W}_{vx}$ are the weights of the three nodes that form a triangle. Those weights are further normalized by the maximum weight in the network:

$$\hat{w} = \frac{w}{max(w)} \quad (2)$$

Figure 2 illustrates how *C* differs among different weighted networks.

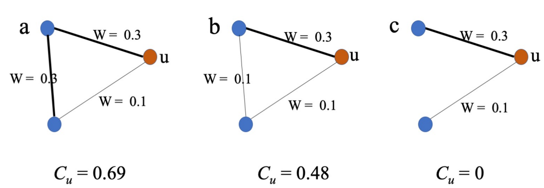

Figure 2: An illustration of the clustering coefficient (C_u) of a node u in three different weighted networks: a, b, and c. Although the numbers of nodes are equal among the networks, the weights of the edges differ. The edge between the two neighbors of the node u has stronger weight in network a than in network b. Therefore, C_u in network a is higher than network b. The lowest C_u is observed in network c because the weight of the edge between the two neighbors equals zero, meaning that the two neighbors are not connected to each other.

Previous research has shown that the clustering coefficient can influence the accessibility of a spoken word (Chan and Vitevitch, 2009, 2010). Thus, words with higher C are harder for people to retrieve than words with lower C. Chan and Vitevitch (2009), in their word recognition task, reported that participants responded more slowly and less accurately to words with high C (e.g. 'dish', 'full', 'lool') than to words with low C (e.g. 'bush', 'wide', 'lick').

To explain the observed patterns, Chan and Vitevitch (2009) proposed that given a constant input, a word with high C will spread activation to its neighbors to a greater extent than a word with low C, as the neighbors of a word with high C are also mutually connected. Consequently, the activation of a word with high C is less distinctive because its neighbors are also activated by reverberating activation from the other neighbors. On the other hand, the activation of a word with low C is much more distinctive because its neighbors do not benefit from other interconnected neighbors. Thus, such a word is accessed faster and is significantly more activated than its neighbors. The proposal was later supported by two simulation studies (Siew, 2019; Vitevitch et al., 2011)

Since most previous studies were done within a monolingual context, it is still not clear whether the network structure can also influence bilingual lexical retrieval. If similar mechanisms are shared by monolinguals and bilinguals in terms of the clustering coefficient of a word, we should also expect its effect on bilingual speech.

2 Present study

The present study aims to examine word-level code-switching mechanisms from a network science perspective. Given that code-switching has been shown to be driven by accessibility, and that the C of a word influences the word's retrieval (Chan and Vitevitch, 2009, 2010), we predict that the clustering coefficient property may capture the likelihood of code-switching of words. Analyzing the corpus with spontaneous bilingual code-switching speech, we hypothesize that a word is code-switched from one language to the other (CS word) if it has a lower clustering coefficient value than its translated equivalent. For instance, in the code-switching sentence '我 很 surprised' (I'm very surprised), we expect the CS word *surprised* to have a lower C than its translated equivalent in Chinese. In addition, word frequency, a traditional indicator of lexical accessibility, will be considered and controlled. This study focuses on English-Chinese balanced bilinguals.

3 Data and network structure

We first introduce the procedures of retrieving CS words and their translated equivalents. Next, we establish semantic representations for words in each language by looking at words as vectors. Finally, semantic networks are constructed based on semantic similarities between words.

3.1 CS words

We used the corpus of Mandarin-English Code-Switching in South-East Asia (Lyu et al., 2015) to retrieve CS words. The corpus provides transcriptions of free conversations from 156 English-Mandarin balanced bilinguals. During conversation, the participants were free to shift between languages, thus naturally producing code-switching speech. Data were pre-processed by excluding non-word markers (e.g., '⟨unk⟩') and words that are in comments. Although Chinese word segmentation had already been applied in the original data, we noticed that some segments still contained more than one word. For example, the segment '我不知道' (i.e., 'I do not know') should have been further segmented into words '我' ('I'), '不' ('not'), and '知道' ('know'). Therefore, we re-applied word segmentation with PKUSEG (Luo et al., 2019), a state-of-the-art segmenter for Chinese words. In total, there are 155,979 sentences being included, with 11,376 unique words in English and 15,428 unique words in Chinese.

As has been described above, a CS word is defined as a word in a different language than its preceding word within a sentence. For example, in the sentence below, the words 'apply', '那个', and 'job' are all considered to be CS words.

刚才 我 不是 跟 你 讲 我 apply 那个 job
"Haven't I told you just now that I applied for that job?"

3.2 Translated equivalents

We used the Princeton WordNet of English (Miller, 1995) and the Chinese Open WordNet (Wang and Bond, 2013) from Open Multilingual WordNet (Bond and Foster, 2013) to translate the CS words. The two WordNets use common word senses. Therefore, if a word in one WordNet shares the same word sense(s) with a word in the other WordNet, the two words are considered translated equivalents of each other (See Figure 3 for an example). Since a word can have multiple translated equivalents, we only kept the best match, i.e. the translation that shared the greatest number of word senses with the target word. If more than one translated equivalent shared the greatest number of word senses, one of the translation was randomly selected.

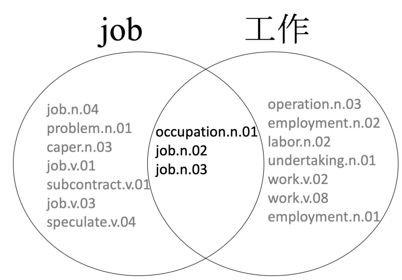

Figure 3: An illustration of how translated equivalents were established through multilingual WordNet. The Chinese word '工作' is a translated equivalent of the English word 'job', as they share three senses.

Due to the size difference between the English and the Chinese WordNets, not all CS words had their translated equivalents. To maintain the quality of the translation, we only kept the CS words whose translated equivalents could be found in the WordNets. This resulted in 3,453 English-Chinese word-translation pairs and 898 Chinese-English word-translation pairs being kept.

3.3 Embeddings

Data We used preprocessed fastText embeddings of English and Chinese (Grave et al., 2018)[1]. Each fastText model contains vectors of 2M unique words. The fastText embedding model is a neural network model that learns vector representations of words from text. Words that appear in similar contexts are closer in vector space. Because semantically-related words tend to exist in similar contexts (e.g., 'king' and 'queen'), fastText can well capture the semantic association between words (Bojanowski et al., 2017). For example, it is able to tell that the semantic association between 'king' and 'man' is analogous to the association between 'queen' and 'woman'. Moreover, fastText outperforms other similar embedding models in semantic representations (Bojanowski et al., 2017; Grave et al., 2018), because it also has embedded fine-grained sublexical information such as morphology in English and characters in Chinese.

[1] https://fasttext.cc/docs/en/crawl-vectors.html

3.4 Semantic networks

A weighted semantic network was built for each language. Because analyzing large weighted semantic networks can be computationally expensive, we constructed the semantic network with the CS words and the translated equivalents rather than all words from the embedding models or the code-switching corpus. Therefore, the nodes in the network of each language represented the CS words and the translated equivalents within that language. The weight of the connection between each of the two words was determined by their cosine similarity calculated from the corresponding word embeddings. In cases where the calculated cosines resulted in negative values, the edges between words will be marked with zero weight.

4 Code-switching analyses

For semantic networks of English and Chinese, the clustering coefficient of each word was calculated. To determine whether the CS words have lower C values than their translated equivalents, we compared the Cs between each CS word and its translated equivalent, in both directions (i.e., English-Chinese and Chinese-English). As word frequency is a classical indicator of lexical accessibility (Gross and Kaushanskaya, 2015; Gollan and Ferreira, 2009; Gollan et al., 2014), we also analyzed and controlled word frequency during the analysis to ensure that the effect of C on code-switching is not a by-product of frequencies.

Word frequency We obtained the frequency per million words of each word from the SUBTLEX databases of English (Brysbaert and New, 2009) and Chinese (Cai and Brysbaert, 2010). The SUBTLEX databases of both languages contain word frequency information from subtitles of film and TV series.

Word-translation pairs were kept only if both the CS word and its translation were found in the SUBTLEX databases. As a result, there were 2,778 English-Chinese word-translation pairs and 721 Chinese-English pairs used in the following analyses. Table 1 provides the sample sizes and examples for each condition. The average frequency per million words is 99.84 ($SD = 703.35$) for English words and 123.95 ($SD = 812.79$) for Chinese words.

	CS words	**translated equivalents**
English-Chinese	2778	2778
e.g., job-工作	job	工作
Chinese-English	721	721
e.g., 那个-that	那个	that

Table 1: Number of words in each condition, and bidirectional word-translation examples. 'job-工作' is a English-Chinese pair where 'job' is the CS word; '工作' is the translated equivalent. '那个-that' is a Chinese-English pair with '那个' being the CS word and 'that' being its translation.

4.1 Data rescaling

Because Cs and frequencies of words in English and Chinese were drawn from different resources, we standardized them so that the values of the two languages can be compared on the same scale.

For Cs, z-scores were used. The Cs for each language were separately standardized as z-scores by subtracting each one from its language's mean C and dividing by its standard deviation. For word frequencies, we used frequency per million words in the SUBTLEX databases (Brysbaert and New, 2009; Cai and Brysbaert, 2010), which were then log-transformed to reduce the skew of raw word frequencies.

4.2 Analysis plan

Controlling covariates To distinguish the effect of Cs from the effect of word frequencies, we first conducted a correlation analysis to see whether Cs and frequencies are correlated. Next, we residualized each of these variables in the statistical analyses to remove the effect of the other variable. Specifically, when testing the effect of C with word frequency being controlled, residualized Cs were calculated from a regression model using frequency as the predictor of C. Similarly, frequencies were residualized with Cs being controlled when testing the frequency effect.

Statistical analysis Parametric and non-parametric tests were used to compare the residualized Cs or residualized word frequencies between CS words and their translated equivalents. For parametric tests, the magnitude of the difference between CS words and their translated equivalents were considered. A mixed-design ANOVA was run using a within-item variable of Switching (whether the word was a CS word or a translation), and a between-item variable of Direction of translation (English-Chinese or Chinese-English). Paired-sample t-tests were used to further analyze the Switching effect in significant interactions. For non-parametric tests, we counted CS word-translation pairs with lower residualized Cs (or residualized word frequencies) for the CS word versus higher residualized Cs (or residualized frequencies) for the switched word. Sign tests were applied to these counts for each translation direction. Table 1 provides the sample size and examples for each condition.

5 Results

5.1 Correlation between Cs and frequencies

A Pearson correlation was conducted between the standardized C and the log transformation of word frequencies. We found a positive correlation between them, $r = .41, p < .001$, indicating that the higher the word's C, the higher the word's frequency.

5.2 Parametric tests

With standardized and residualized C as the dependent variable, a mixed-design ANOVA showed a significant interaction between Switching and Direction of translation, $F(1, 3497) = 11.25, p < .001$. Paired-sample t-tests found that Cs of CS words were significantly lower than their translated equivalents for both the English-Chinese translation direction ($t(2778) = -10.47, p < .001, d = -.20$) and the Chinese-English translation direction ($t(721) = -7.47, p < .001, d = -.28$), although the difference for the Chinese-English was greater. The results indicate that, with word frequency controlled, the CS words have lower Cs than their translated equivalents.

With residualized log-transformed frequency as the dependent variable, the interaction between Switching and Direction was also significant, $F(1, 3497) = 149.93, p < .001$. Unlike the results for Cs, paired-sample t-tests showed significantly higher word frequencies in CS words than their translated equivalents for both the English-Chinese direction ($t(2778) = 13.31, p < .001, d = .25$) and the Chinese-English direction ($t(721) = 17.90, p < .001, d = .67$). The difference for the Chinese-English direction was greater than for Chinese-English, a pattern that is opposite to that of the Cs.

To conclude, CS words have lower Cs but higher frequencies than their translated equivalents, despite the positive correlation between Cs and frequencies. The opposite effects of Cs and frequencies on code-switching are shown in Figure 4.

5.3 Non-parametric tests

For both English-Chinese and Chinese-English translation directions, sign tests indicate that there were significantly more CS word-translation pairs in which the CS word had lower C (standardized and residualized) but higher word frequency (log-transformed and residualized) than its translated equivalent, which is consistent with the parametric test results. Table 2 reports the detailed statistical output.

6 Testing alternatives

In this section, we test a potential alternative to the proposed hypothesis that the C of words may affect bilinguals' code-switching speech.

The original distributions of the raw C values for the two languages, as was shown in Figure 5, were largely disparate. While this is possibly an artifact due to the Cs for the two languages being derived from separate networks, an alternative explanation is that it arises from a cross-linguistic difference of the words' Cs themselves, such that Chinese words tend to have higher C than English words. If there were an effect of such a cross-linguistic difference, data transformations might blur this effect; hence the results after the transformations might be questionable.

We indirectly test our argument about the effect of C against the alternative by comparing the Cs of the CS words and the translated equivalents $within$ each language, although now the CS words and the translated equivalents are two independent groups (i.e., unpaired). If our hypothesis is correct, that bilinguals tend to switch to the other language when the C of the word is lower than its translation, then we should also expect that, within the same language, the words that are CS words should have lower Cs than the words that are translated equivalents of words in the other language.

Analysis The words analyzed in each language were either the CS words or the words that were translated equivalents of the CS words from the other language. For each language, an independent-samples t-test was used to compare the difference

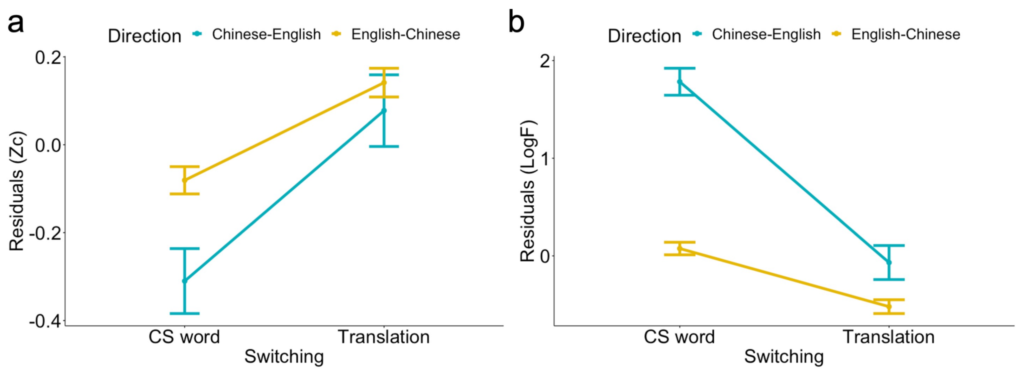

Figure 4: Means for Switching by Direction of translation for (a) standardized and residualized C, residuals (Z_c), and (b) log-transformed and residualized frequency, residuals ($LogF$). The error bars represent the 95% confidence intervals.

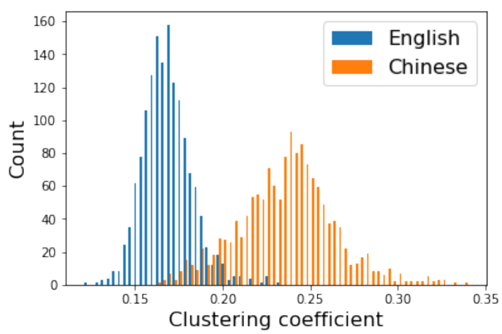

Figure 5: The frequency distributions of raw C for English and Chinese words.

of C (standardized and residualized) between these two groups of words.

Results The CS words had lower C than the words that were translated equivalents, in both languages. In Chinese, the Cs of CS words ($M = -.31, SD = 1.01$) were significantly lower than the Cs of words that are translated equivalents ($M = .14, SD = .88$), $t(3497) = 11.90, p < .001, d = .50$. In English, the CS words ($M = -.08, SD = .84$) also had significantly lower Cs than the translated equivalents ($M = .08, SD = 1.11$), $t(3479) = 4.21, p < .001, d = .18$. After controlling for the potential cross-linguistic differences, the results are still compatible with our argument that bilinguals tend to switch to the other language when the C of the CS word is lower than its translation.

7 Discussion and conclusion

We have demonstrated that the CS words tend to have lower Cs than their translated equivalents. The results are consistent in that the observed pattern that the CS word has lower C than its translated equivalent was supported by both parametric and non-parametric tests. With word frequencies controlled, we also ruled out the possibility that the influence of C is merely a byproduct of a word frequency effect.

More interestingly, we observed opposite effects of the C and the word frequency on code-switching, despite the positive correlation between the C and the frequencies. Unlike C, the CS words tend to have higher word frequencies than their translated equivalents that were being replaced. On the one hand, our findings support previous studies that words with high frequencies were more likely to be chosen during free code-switching speech production (Gollan and Ferreira, 2009; Gross and Kaushanskaya, 2015; Gollan et al., 2014). On the other hand, the disparity between the effects of C and word frequency on code-switching indicates the previously unnoticed properties of C on code-switching. That is to say, C, as an important metric in network science, makes its own contribution to language processing. The findings further underscore the importance of studying words in a global context that incorporates the interconnections and the interactions between words (Karuza et al., 2016).

The findings echo previous research that words with lower C were more easily retrieved than words with higher C in language production (Chan and

Difference pattern	English-Chinese	Chinese-English
Count (residualized Z_c)		
$C_{CSword} < C_{translation}$	1658	433
$C_{CSword} > C_{translation}$	1121	288
χ^2	104.19***	29.16***
Count (residualized $LogF$)		
$F_{CSword} < F_{translation}$	1091	170
$F_{CSword} > F_{translation}$	1678	551
χ^2	127.87***	201.33***

Table 2: Sign tests for Cs and word frequencies. 'English-Chinese' and 'Chinese-English' denote the two translation directions. Z_c represents the standardized C, whereas $LogF$ represents the log transformed word frequency. $C_{CSword} < C_{translation}$ represents the number of CS word-translation pairs in which the CS word had lower C (standardized and residualized) than its translated equivalent, whereas $C_{CSword} > C_{translation}$ represents the opposite pattern. $F_{CSword} < F_{translation}$ represents the number of CS word-translation pairs in which the CS word had lower word frequency (log transformed and residualized) than its translated equivalent, and $F_{CSword} > F_{translation}$ represents the opposite pattern.

Vitevitch, 2009, 2010). More specifically, it is likely that a word with lower C 'stands out' among its neighbors, whereas its translated equivalent with higher C is likely to be less distinguishable from its interconnected neighbors. Therefore, bilinguals can retrieve the word with lower C more easily. The cost of doing that, however, is the necessity of switching between two languages. The present study further complements, from a semantic and a bilingual perspective, previous work on phonological networks. To deepen the understanding of C in semantic networks, future studies could apply experimental paradigms used by (Chan and Vitevitch, 2009). To explore the influence of C in bilingual speech, future work could extend the experimental tests to words from different languages with bilinguals participating in the task.

Our study provides a novel approach to investigating code-switching in speech, which further sheds light on the possibility that bilingual word retrieval is influenced by the interconnection and the interaction between words. Nevertheless, the pattern observed in our study, where a CS word tends to have lower C than its translated equivalent, does not directly address the proposed mechanism that the activation of the lower C word 'stands out' among its neighbors and therefore is selected over its translated equivalent. Future work should examine the underlying mechanism, through computational modeling, experimental studies, to name a few.

We have only looked at English-Chinese bilinguals. If the influence of C on code-switching holds true, it should also be observed in bilinguals speaking different pairs of languages. Future studies should test the generalizability of this claim. The quality of the translated equivalents could be evaluated and enhanced in the future by using a professional bilingual dictionary, asking bilinguals to evaluate the translations, and considering word context. In the present study, although we used WordNet to find the best matching translation in general, we could not guarantee that those translations were correct given the context in the sentence. One possible way of tackling this problem is to apply sense-specific word representations (Ettinger et al., 2016; Upadhyay et al., 2017) to represent and match the particular senses of words in the two languages.

The semantic representation in our study was built for each language separately. This is because the separate semantic models allow us to assess the properties of the words being replaced (i.e. translated equivalents of the CS words). In a code-switching corpus, on the other hand, such properties are likely to be hidden. However, studies have shown that the two lexicons of a bilingual speakers share some representation, at least at the semantic level (Francis, 1999), and that a bilingual is not necessarily the sum of two monolinguals (Grosjean, 1989). Consequently, it is likely that the representations of two languages in bilingual speakers are different from the representations of two monolingual languages. To overcome this limitation, future studies could test the effect of a word's C on code-switching using human-subject experiments. A more extended and comprehensive bilingual corpus would also be beneficial as it could reveal information about both the CS words and their translated equivalents that is not present in monolingual corpora.

References

Utsab Barman, Amitava Das, Joachim Wagner, and Jennifer Foster. 2014. Code mixing: A challenge for language identification in the language of social media. In *Proceedings of the First Workshop on Computational Approaches to Code Switching*, pages 13–23.

Leslie M Beebe and Howard Giles. 1984. Speech-accommodation theories: A discussion in terms of second-language acquisition. *International Journal of the Sociology of Language*, 1984(46):5–32.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual WordNet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1352–1362.

M. Broersma. 2009. Triggered codeswitching between cognate languages. *Bilingualism: Language and Cognition*, 12:447–462.

Angela de Bruin, Arthur G Samuel, and Jon Andoni Duñabeitia. 2018. Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language*, 103:28–43.

Marc Brysbaert and Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4):977–990.

Barbara E Bullock and Almeida Jacqueline Ed Toribio. 2009. *The Cambridge handbook of linguistic code-switching.* Cambridge University Press.

Qing Cai and Marc Brysbaert. 2010. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PloS One*, 5(6):e10729.

Kit Ying Chan and Michael S Vitevitch. 2009. The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 35(6):1934–1949.

Kit Ying Chan and Michael S Vitevitch. 2010. Network structure influences speech production. *Cognitive Science*, 34(4):685–697.

Maria Carmen Parafita Couto and Marianne Gullberg. 2019. Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism*, 23(2):695–714.

Allyson Ettinger, Philip Resnik, and Marine Carpuat. 2016. Retrofitting sense-specific word vectors using parallel text. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1378–1383.

Wendy S Francis. 1999. Cognitive integration of language and memory in bilinguals: Semantic representation. *Psychological Bulletin*, 125(2):193.

Matthew Goldrick and Brenda Rapp. 2007. Lexical and post-lexical phonological representations in spoken production. *Cognition*, 102(2):219–260.

Tamar H Gollan and Victor S Ferreira. 2009. Should I stay or should I switch? A cost–benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3):640.

Tamar H Gollan, Daniel Kleinman, and Christina E Wierenga. 2014. What's easier: Doing what you want, or being told what to do? Cued versus voluntary language and task switching. *Journal of Experimental Psychology: General*, 143(6):2167.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893*.

François Grosjean. 1989. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36(1):3–15.

Megan Gross and Margarita Kaushanskaya. 2015. Voluntary language switching in English–Spanish bilingual children. *Journal of Cognitive Psychology*, 27(8):992–1013.

Thomas T Hills, Mounir Maouene, Josita Maouene, Adam Sheya, and Linda Smith. 2009. Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science*, 20(6):729–739.

Elisabeth A Karuza, Sharon L Thompson-Schill, and Danielle S Bassett. 2016. Local patterns to global architectures: influences of network topology on human learning. *Trends in Cognitive Sciences*, 20(8):629–640.

Daniel Kleinman and Tamar H Gollan. 2016. Speaking two languages for the price of one: Bypassing language control mechanisms via accessibility-driven switches. *Psychological Science*, 27(5):700–714.

Ping Li. 1996. Spoken word recognition of code-switched words by Chinese–English bilinguals. *Journal of Memory and Language*, 35(6):757–774.

Ruixuan Luo, Jingjing Xu, Yi Zhang, Xuancheng Ren, and Xu Sun. 2019. PKUSEG: A toolkit for multi-domain Chinese word segmentation. *arXiv preprint arXiv:1906.11455*.

Dau-Cheng Lyu, Ren-Yuan Lyu, Yuang-chin Chiang, and Chun-Nan Hsu. 2006. Speech recognition on code-switching among the Chinese dialects. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 1, pages I–I.

Dau-Cheng Lyu, Tien-Ping Tan, Eng-Siong Chng, and Haizhou Li. 2015. Mandarin–English code-switching speech corpus in South-East Asia: SEAME. *Language Resources and Evaluation*, 49(3):581–600.

Viorica Marian. 2009. Language interaction as a window into bilingual cognitive architecture. In Ludmila Isurin, Don Winford, and Kees De Bot, editors, *Multidisciplinary Approaches to Code Switching*, pages 161–185. John Benjamins.

George A Miller. 1995. WordNet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Carol Myers-Scotton. 1993. Common and uncommon ground: Social and structural factors in codeswitching. *Language in Society*, 22(4):475–503.

Jukka-Pekka Onnela, Jari Saramäki, János Kertész, and Kimmo Kaski. 2005. Intensity and coherence of motifs in weighted complex networks. *Physical Review E*, 71(6):065103.

Shana Poplack. 2000. Sometimes I'll start a sentence in spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *The Bilingualism Reader*, 18(2):221–256.

Beatrice Santorini and Shahrzad Mahootian. 1995. Codeswitching and the syntactic status of adnominal adjectives. *Lingua*, 96(1):1–27.

Cynthia SQ Siew. 2019. Spreadr: An R package to simulate spreading activation in a network. *Behavior Research Methods*, 51(2):910–929.

Ann E Sizemore, Elisabeth A Karuza, Chad Giusti, and Danielle S Bassett. 2018. Knowledge gaps in the early growth of semantic feature networks. *Nature Human Behaviour*, 2(9):682–692.

Mark Steyvers and Joshua B Tenenbaum. 2005. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78.

Holly L Storkel, Jonna Armbrüster, and Tiffany P Hogan. 2006. Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, 49(6):1175–1192.

Shyam Upadhyay, Kai-Wei Chang, Matt Taddy, Adam Kalai, and James Zou. 2017. Beyond bilingual: Multi-sense word embeddings using multilingual context. *arXiv preprint arXiv:1706.08160*.

Michael S Vitevitch, Kit Ying Chan, and Steven Roodenrys. 2012. Complex network structure influences processing in long-term and short-term memory. *Journal of Memory and Language*, 67(1):30–44.

Michael S Vitevitch, Gunes Ercal, and Bhargav Adagarla. 2011. Simulating retrieval from a highly clustered network: Implications for spoken word recognition. *Frontiers in Psychology*, 2:369.

Shan Wang and Francis Bond. 2013. Building the Chinese Open WordNet (COW): Starting from core synsets. In *Proceedings of the 11th Workshop on Asian Language Resources*, pages 10–18.

Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.

Mark Yates. 2013. How the clustering of phonological neighbors affects visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5):1649–1656.