Large-scale Network Analyses Reveal Cross-Language Differences in Semantic Structures: A Comparative Study

Qihui Xu (q.xu@bcbl.eu)
Basque Center on Cognition, Brain and Language (BCBL)
Donostia-San Sebastián, Spain

Yingying Peng (yingying1.peng@connect.polyu.hk)
The Hong Kong Polytechnic University
Hong Kong, China

Ping Li (ping2.li@polyu.edu.hk)
The Hong Kong Polytechnic University
Hong Kong, China

Abstract

English and Mandarin Chinese are two distinct languages in many aspects, such as orthography and morphology. Previous network analyses show strong clustering coefficients (C) on English semantic networks, revealing the interconnectedness of semantic representations between words. However, it is not clear whether such semantic representation properties are language specific or general, and whether the linguistic-feature difference (e.g., subword components such as orthography and morphology) may affect the lexico-semantic structure. Here, we compared Cs of words in English and Mandarin semantic networks based on a) feature norms empirically derived from human subjects and b) distributed semantic information of text retrieved by word embedding models. We consistently observed higher Cs of Mandarin words than English words, especially when the semantic network considers subword features. Linear regressions suggested that the subword components' semantic properties in Mandarin, but not in English, could significantly and positively predict the C of words in semantic networks. The results indicate an important role of language-specific properties in lexico-semantic structures and imply the diversity of human language processing.

Keywords: Network science; Semantic networks; Cross-linguistic comparison; Feature norms; Word embeddings; Computational modeling

Introduction

Network science benefits the understanding of human lexical-semantic representation (Steyvers and Tenenbaum, 2005; Borge-Holthoefer and Arenas, 2010) by amassing a vast amount of words traditional methods typically fall short of (e.g., behavioral studies) and providing global insight into the interconnections, interactions and higher-order organization of words (Karuza et al., 2016). Normally, a semantic network is constructed with words as its nodes and semantic relations/similarity among words as edges. The global semantic system of words has been shown to affect many aspects of language processing (Hills et al., 2009a, 2009b; Xu et al., 2021) and has revealed unique features such as strong interconnectedness of semantic representations between words. However, most previous studies have focused on a single language. It is unclear whether such semantic representation properties are language specific or general, and whether the cross-linguistic difference, such as orthography and morphology, may be associated with lexico-semantic structure.

Subword features in English and Mandarin Mandarin Chinese and English are two distinct languages in orthography and morphology, among other linguistics features. In Mandarin Chinese script, four levels of structural complexity are involved in a word: stroke, radical, character, and word (see Figure 1; Yeh et al., 2017). The basic units are strokes. Strokes could be combined as fixed sets to form radicals that typically carry semantic or phonetic information about the character. Approximately 96% of commonly used Mandarin Chinese characters are constituted by semantic and phonetic radicals (Li and Kang, 1993; Hsiao et al., 2007). Semantic

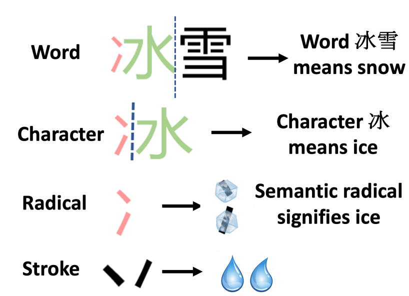


Figure 1: A demonstration of four levels in a Mandarin Chinese word. Note that a character may also be a word by itself

radicals generally reflect the semantic categories of Chinese characters and words (Huang and Hsieh, 2015) and facilitate lexical-semantic processing (Feldman and Siok, 1999; Ding et al., 2004). For example, Mandarin Chinese characters with higher semantic transparency (i.e., the degree to which the meaning of the character is directly related to the meaning of its semantic radical) are processed more accurately and faster (Hsiao et al., 2007). Moreover, Chinese characters with semantic radicals that possess a larger family size (i.e., number of characters sharing the same semantic radicals; Feldman and Siok, 1997) are recognized faster than those with a smaller semantic radical family size. Similarly, Studies in English found that the morphological family size could influence lexical-semantic recognition and processing (Baayen et al., 2006). However, with letters as fundamental units in its writing system, English does not possess semantic subword

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

components as prevalent as semantic radicals in Mandarin words (Shen and Ke, 2007). Will the ubiquitous semantic radicals in mandarin Chinese influence Mandarin speakers' semantic knowledge organization and representation? We infer that Mandarin Chinese characters/words with the same semantic radicals might be "stored" in the mental lexicon in a more clustered manner due to a higher degree of similarity in form.

Clustering coefficient and lexical-semantic processing Clustering coefficient (C) of a word in a semantic network quantifies the extent to which the word's semantically similar words are also similar to each other (Steyvers and Tenenbaum, 2005). It reflects how clustered (i.e., grouped together) the semantic representations are for the word and its semantically similar words. The C of a node in a weighted network is calculated by taking the sum of the geometric average of the edge weights of that node (Onnela et al., 2005):

$$C_u = \frac{2}{\deg(u)\,(\deg(u)-1)} \sum_{vw} (\hat{W}_{uv}\hat{W}_{uw}\hat{W}_{vw})^{1/3} \qquad (1)$$

where $deg(u)$ represents the number of edges to which a given node is connected (D). The weights during the calculation of C are normalized via scaling the weights by the largest weight in the network (Onnela et al., 2005) using the formula $\widetilde{W_{ij}} = W_{ij}/max(W_{ij})$. Figure 2 presented an illustration of the C of a node in different weighted networks.

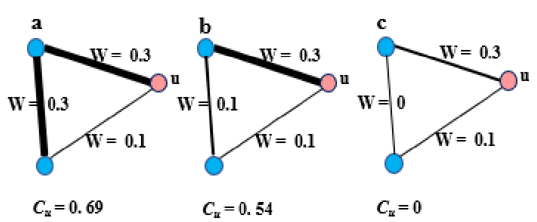


Figure 2: An illustration of the C of a node u in three different weighted networks: a, b, and c. The three networks have an equal number of nodes but differ in edge weights.

C is a commonly studied metric in semantic networks. It has been shown to influence the inhibition and facilitation of lexical processing via the spreading activation between a word and its neighbors (Vitevitch et al., 2011; Vitevitch and Luce, 2016; Xu et al., 2021). Given its implication in lexical-semantic processing, we focus on the cross-linguistic comparison on C. If rich semantic information of Mandarin Chinese radicals affects the semantic structure, then the semantic representation of Mandarin words is likely to be more clustered; thus words in a Mandarin semantic network may have higher C than in English.

Present study

In this study, we have two main objectives. First, we aim to compare the lexical-semantic structure across languages, focusing on English and Mandarin Chinese. We analyze the clustering coefficient (C) of semantic networks built using feature norms, and word embedding models including both word and subword information. Considering that subword components influence word semantics and Mandarin radicals carry rich semantic information (Feldman and Siok, 1997), we hypothesize higher C values in Mandarin networks than English. Furthermore, we expect the cross-linguistic difference in C to be more pronounced in embedding networks with subword-level information, further highlighting the role of subword components.

Second, we aim to directly investigate the role of subword components in semantic structures and examine potential cross-linguistic differences. In view of previous findings on semantic transparency's influence on lexical-semantic representation and processing, we introduce a new metric, semantic consistency (SemC), to quantify the subword component's impact. SemC is computed based on the average semantic similarities between words sharing the same subword component. Figure 3 illustrates SemC computation. In step 1 (Figure 3a), we extracted subword components from individual English and Chinese words (e.g., "second" and "hand" for "secondhand" in English, and "扌" and "丁" for "打" in Chinese). [1].

In step 2, we computed each subword component's SemC by measuring the average semantic similarities between words sharing the same components (Figure 3b). A higher average semantic similarity suggests a more consistent semantic representation for that component. For example, the Mandarin radical "扌" may have higher SemC, as words containing it, like "打" (hit), "推" (push), and "捏" (pinch), are semantically similar. Similarly, "hand" in English may have high SemC, as words like "handicraft", "handkerchief", "secondhand", and "handicap" are semantically related. Lastly, we computed each word's SemC (henceforth SemC-word) by averaging[2] the SemCs of its subword components (Figure 3c). We then used the calculated SemC to predict C. As semantic subword components are more prevalent in Mandarin than in English (Li and Kang, 1993; Shen and Ke, 2007), we hypothesize that subword components' semantic properties significantly predict word C in Mandarin semantic networks, but not in English.

Network Structure Comparison Across Languages

For each language, we built semantic networks using feature norms (Deng et al., 2021; Devereux et al., 2014) and word embeddings (Bojanowski et al., 2017; Cao et al., 2018; Mikolov et al., 2013). These resources are comparable across languages, and Andrews et al. (2009) demonstrated that the

[1]For English, this was done with a Python package polyglot that offers trained morfessor models (F = 0.76; Virpioja et al., 2013) to generate morphemes from words. https://polyglot.readthedocs.io/en/latest/MorphologicalAnalysis.html. For Mandarin, this was done with an online dictionary https://tool.httpcn.com/zi/

[2]We also tried the maximum SemC and found that the maximum and mean measurements were strongly correlated, $R_{spearman} > 0.9$

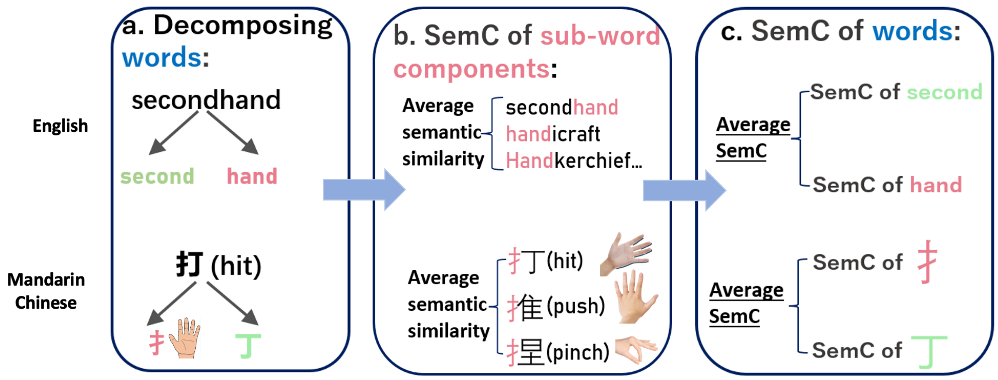


Figure 3: An illustration of the calculation of SemC for English and Mandarin Chinese words.

two approaches together offer better insights into word semantic representations. We did not use word association (Nelson et al., 2004, another common resource for constructing semantic networks, due to the lack of comparable Chinese word association norms.

Semantic feature norms are conceptually and experientially-based (e.g., "banana is a fruit"), providing a transparent semantic similarity through overlapping features among words (Mirman and Magnuson, 2008. Their use in constructing semantic networks has been validated in previous studies (Hills et al., 2009b; Peters and Borovsky, 2019).

Embedding models like word2vec are context-based, learning word vector representations from text corpora. Semantically-related words in similar contexts are closer in vector space, making these models suitable for capturing semantic associations (e.g., "king" and "queen"). Embedding-derived semantic networks have demonstrated psychological plausibility through network metrics, including C (Kajic and Eliasmith, 2018; Steyvers and Tenenbaum, 2005; Utsumi, 2015).

Feature networks

Data To ensure dataset comparability between languages, we used the Centre for Speech, Language and the Brain concept property norms for English (Devereux et al., 2014) and the Chinese Conceptual Semantic Feature Dataset for Chinese (Deng et al., 2021). The tasks were similar across both languages. Participants saw a concept word in each trial and provided at least five features using provided relation words (e.g., "banana is a fruit," where "banana" is the concept, "is" the relation, and "a fruit" is the feature). For both datasets, each concept received responses from at least 30 participants.

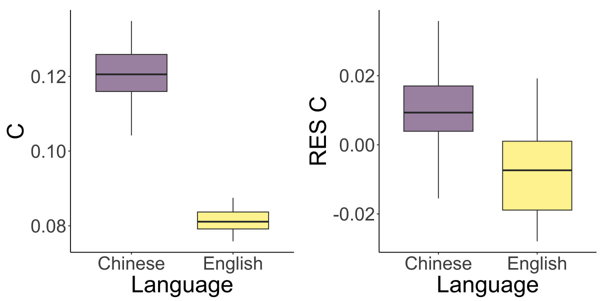
The selection criteria for concept norms were also similar in both languages, prioritizing words with higher concreteness. Mandarin data included all concepts translated from English data. However, the sample sizes differed: 638 concept words for English and 1,410 for Mandarin. We controlled for sample size when constructing and analyzing semantic networks (details provided below).

Determining semantic relations The semantic relations between each of the two concept words were determined by a semantic similarity measurement that both Devereux et al. (2014) and Deng et al. (2021) adopted. As semantically similar words tend to share features, concept words×feature matrices were constructed to represent semantic information of words. Each word has its semantic vector decomposed from its feature representations. Cosine similarity was then used to measure the vector distance between every two words. The greater the cosine between two words - suggesting more shared features - the higher the semantic similarity is between the two words.

Network construction For each language, we constructed a weighted semantic network with the concept words as nodes, and their semantic similarity as weights of the network connections.

Because the large difference in the number of concept words between the datasets of the two languages may confound the cross-linguistic comparison of C, we applied a without-replacement bootstrap method (see Bertail, 1997) to address the issue. For each language, we bootstrapped 100 sub-networks from the original semantic network; each sub-network contains 500 nodes randomly sampled from the original network and the weighted connections between the sampled nodes. Subsequential statistical comparisons were performed on the bootstrapped sub-networks that had identical sizes.

Analysis An average C was computed for each bootstrapped sub-network by averaging the Cs of all nodes. The difference of Cs between the two languages' sub-networks was then tested using an independent-sample t-test. Because D may often confound C, we also obtained average Ds for the sub-networks. To rule out any interference from D, residualized average C (RES C) was calculated from a regression model using average D of each sub-network as a predictor of



(a) (b)

Figure 4: *C*s (a) and RES *C*s (b) of the feature networks in each language

average *C* (see Garcıa et al., 2020 for the validity of residualization). We also statistically compared networks' RES *C*s between the two languages[3].

Results Normality has been assumed in the data both with and without residualization. An independent-sample t-test showed that words in Mandarin semantic networks have greater average *C* ($M = 0.23, SD = 0.02$) than words in English semantic networks ($M = 0.17, SD = 0.01$), $t(196.47) = 12.79, p < .001, d = 3.30$; Figure 4a). A test on RES *C* with *D* controlled showed the same pattern (Figure 4b), $t(198) = 5.23, p < .001, d = 1.35$.

Embedding networks

Training data We downloaded cleaned Wikipedia dumps of English and Mandarin (Yang et al., 2018; Yu et al., 2017) as the training data for embeddings. Both datasets involved a similar filtering process and were retrieved from the archive of the same year (2017). As the two data differ greatly in word tokens (1,615,338,431 for English as compared to 164,998,611 for Mandarin) and unique word types (3,194,435 for Mandarin as compared to 163,329 for English), the larger data was down-sampled to match the size of the smaller data. Because the type-token ratio is different between the two datasets, we performed two versions of downsamples - a matched-type version and a matched-token version. When downsampling, we randomly sampled documents from the larger data, scrambled the order of the documents, processed documents one-by-one, sequentially included sentences from a document being processed while counting the size of the included data every time after processing a sentence, and stopped until the total included size approximately matched (may still differ slightly as each sentence was a processing unit) the smaller data. This approach yielded 164,998,612 tokens in the down-sampled English data of the matched-token version as compared to 164,998,611 for Mandarin, and 163,328 types in the down-sampled Mandarin data of the matched-type version as compared to 163,329 for English.

Word frequency We used the word frequency for selecting words of the networks. SUBTLEX was used (Brysbaert et al., 2011; Cai and Brysbaert, 2010), a large-scale database with word frequency information gathered from film and television subtitles. SUBTLEX of English and Mandarin are comparable in terms of data collection and processing.

Training semantic embeddings The semantic relations between words were determined by semantic embeddings trained on large-scale text data, which were represented at two different levels: the word and the subword levels. Word2vec (Mikolov et al., 2013) was used as the word-level embedding model.

The subword-level embeddings represent fine-grained subword features of words. The smallest subword unit in written English is the letter; in written Mandarin Chinese, it is the stroke. To capture the subword information in semantic representations, we constructed embeddings based on a letter n-gram fastText of English (Bojanowski et al., 2017) and a stroke n-gram cw2vec of Chinese[4] (Cao et al., 2018).

FastText (Bojanowski et al., 2017) is similar to word2vec, except that it learns vector representations not from words, but rather built by the sum of the letter n-grams contained in this word. For example, the vector space of the word ”running” in a trigram fastText model will be a sum of the vectors of its character trigrams ”run”, ”unn”, ”nni”, ”nin”, and ”ing”. Cw2vec of Chinese (Cao et al., 2018) is analogous to fastText of English, except that the n-grams are based on strokes rather than letters. For example, the vector space of the Chinese word ”打” (hit) in a trigram cw2vec model consists of the vectors of its stroke trigrams ”一 亅 ノ ”, ” 亅 ノ 一”, and ” ノ 一 亅 ”. In the actual model training, the n for the n-grams often varies within a range to better capture different sizes of subword components; that is, morphemes in English or, equivalently, radicals in Mandarin (Bojanowski et al., 2017; Cao et al., 2018). With extra subword knowledge, such as morphemes ”ing” in the English word ”running” and radicals ”一 亅 ノ” (i.e. ” 扌”, a radical with semantic meanings of hand-related actions) in the Chinese word ”打”, both fastText of English and cw2vec of Chinese have been shown to better capture words' meanings relative to word2vec models (Bojanowski et al., 2017; Cao et al., 2018).

Embeddings were trained on each language's matched-type/token versions of the Wikipedia dump. During the training, we included parameter settings set by previous studies (Bojanowski et al., 2017; Cao et al., 2018; Mikolov et al., 2013). The dimensionality was 300, the window size was 5, and the n-gram size ranged between 3 and 6 for all of the

[3]Previous studies utilizing unweighted networks normally compare the *C* of the network with that of a random reference network with the same size and degree. However, for weighted networks, constructing a corresponding random reference network involves many other considerations. Thus, instead of constructing random weighted networks, we controlled the node size and degree.

[4]The reason for not including a fastText of Chinese is because the fastText of Chinese takes character n-gram and is fundamentally different from that of English. A preliminary analysis shows that the network built on the Chinese fastText also reveals higher *C*s of words than the English network, which is consistent with the results reported below.

embedding models.

Network construction For each language, we constructed semantic networks using the top 5,000 most frequent words as the nodes and their semantic relations as connections. Therefore, all embedding networks are equal in size. Connections were weighted using the respective embedding model's cosine similarity, a measurement of the cosine of the angle between two vectors. The metric has been widely used in the literature to compute semantic similarities (Bojanowski et al., 2017; Cao et al., 2018; Mikolov et al., 2013).

Analysis As the sizes of the embedding networks are identical between the two languages, here we analyzed the whole embedding network rather than bootstrapped sub-networks. For each network, the C of individual words was calculated. The cross-linguistic comparison of C was then made through a mixed-model ANOVA, with language (English vs. Mandarin) as a between-subject independent variable and representation level (word vs. subword) as a within-subject independent variable. We also tested C with D controlled through residualization (i.e., RES C). We tested the matched-type and the matched-token versions separately.

Results Because our data did not meet the normality of residuals assumption, we used robust statistical methods (Mair and Wilcox, 2020) which are shown to be robust to violations of assumptions of statistical tests such as ANOVA and t-test.

Matched token. Figure 5 illustrated the results. Robust ANOVA showed a significant interaction effect between language and representation level on C, $F(1, 3121.64) = 69585.8, p < .001$. For the word-level embedding network, Chinese words ($M = 0.18, SD = 0.02$) had significantly higher C than English words ($M = 0.14, SD = 0.02$), $t(4752.84) = 100.38, p < .001, d = 2.25$. Differences in Cs between Chinese words and English words were magnified in the subword-level network, $t(3869.39) = 288.93, p < .001, d = 6.45$, suggesting the effect of subword information when evaluating differences in C across languages.

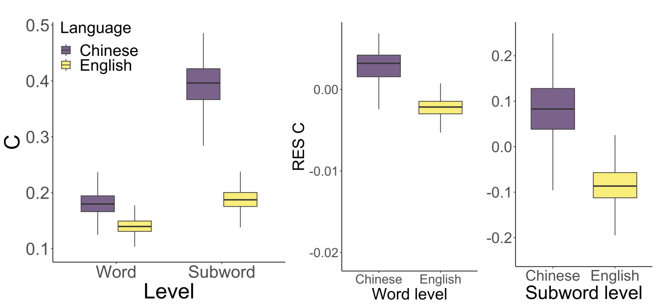


Figure 5: Cs and RES Cs of the embedding networks in each language and representation level. To present the data clearly, the range of the y-axis in the word-level RES C is smaller than in the subword-level RES C graph.

Similarly, analyzing RES C (Figure 5) showed a) a significant interaction effect between language and representation level, $F(1, 4587.61) = 16413.09, p < .001$, b) significantly higher RES Cs of Mandarin words than those of English words in the word-level networks, $t(4263.17) = 134.38, d = 2.10$, and c) larger difference in the subword-level networks of RES C between the two languages, $t(4620.2) = 138.83, p < .001, d = 3.01$.

Matched type. The results of the matched-type embedding networks are consistent with the matched-token networks. Both C and RES C of words in the Mandarin networks were significantly higher than in English in the word-level networks ($t(4752.84) = 146.24, p < .001, d = 2.32$ for C; $t(4263.17) = 87.64, p < .001, d = 1.85$ for RES C) and the subword-level networks ($t(3869.39) = 174.36, p < .001, d = 4.35$ for C; $t(4620.2) = 122.14, p < .001, d = 2.01$).

Semantic Consistency of Subword Components and Network Structure

Measuring semantic consistency SemC calculation has been introduced above (Figure 3). To determine the semantic similarity, we used cosine similarities obtained from the feature norms (Deng et al., 2021; Devereux et al., 2014) and the subword-level embeddings (fastText for English Bojanowski et al., 2017 and cw2vec for Mandarin Cao et al., 2018). Analyses will be cross-validated with those two separate cosines to ensure that any observed relation between SemC and networks' C is not due to an artifact that the same cosines were used for constructing networks and calculating SemC.

Analysis In each network, we computed a word2vec-based SemC-word and a feature-based SemC-word for each word. Next, we used linear regression models with SemC-word as a predictor and the word's C of a semantic network as the outcome. Word frequency[5](Brysbaert et al., 2011; Cai and Brysbaert, 2010), one most commonly studied lexical property, was also included as a covariate. As some words solely appear in feature or embedding networks, when using SemC estimated from one source to analyze networks from the other source, only common words of the two networks were included in regressions. Table 1 listed the number of words being analyzed on different conditions.

Results Table 1 reported the results. For all the Mandarin networks, SemC-word was a significant, positive predictor of the C of words and is independent of and stronger than the prediction of word frequency. For example, embedding-based SemC-word was a significant, positive predictor of words' C in the Mandarin feature network, $t = 12.15, p < .001$. Including SemC-word and frequency as predictors significantly improved the model fit ($R^2 = 0.14$) as compared to a model with word frequency only ($R^2 = 0.03$), $F_{change}(1, 1100) = 147.72, p < .001$. SemC-word explained

[5]Other lexical properties such as concreteness may also affect semantic organizations of the lexicon (Fliessbach et al., 2006). However, there are no adequate norms available currently in Mandarin. Therefore, we leave other covariates for future explorations

Table 1: Prediction of SemC-word of the C of words in linear regressions. Word frequency was controlled. "Words" represent the number of words included in the regression. b is the unstandardized coefficient. R^2_{change} indicates the additional variation in C explained by adding SemC-word as a predictor as compared to a model where word frequency was the only predictor. Embedding networks here are the matched-token version, but the matched-type version revealed consistent results. $^{***}p < .001$; $^{**}p < .01$.

Language	Network	SemC source	Words	b	R^2_{change}
English	Feature network	Feature norms	420	0.07**	0.03**
		Embedding	348	0.06	0.00
	Word-level embedding network	Feature norms	216	0.00	0.00
		Embedding	5000	-0.00	0.00
	Subword-level embedding network	Feature norms	216	-0.01	0.02
		Embedding	5000	0.02***	0.01***
Mandarin	Feature network	Feature norms	1195	0.29***	0.12***
		Embedding	1102	0.28***	0.11***
	Word-level embedding network	Feature norms	185	0.07***	0.07***
		Embedding	5000	0.08**	0.03**
	Subword-level embedding network	Feature norms	185	0.29***	0.09***
		Embedding	5000	0.22***	0.08***

an additional 11% of variation in words' C. In contrast, the prediction of SemC-word of words' C in the English semantic networks was either insignificant or weak.

Discussion

The present study involves a systematic comparison between Chinese and English semantic structures and reveals higher-order differences between the two languages. By examining semantic networks based on empirically-derived feature production and computationally-derived text data, our work consistently revealed that words in Mandarin semantic networks have higher C than words in English semantic networks. By manipulating representation level (e.g., word vs. subword levels) in the embedding models, we found that models representing subword components showed a greater cross-linguistic difference in C. These findings may indicate that semantic representations of words in Mandarin Chinese are more interconnected than the semantic representations of words in English at a large-scale structural level.

By quantifying semantic transparency with our SemC measurement, we found that SemC significantly predicted C in Mandarin, independent of and stronger than word frequency. In English, SemC's prediction of C was either weak or insignificant. These results highlight the unique role of Mandarin-specific subword components in shaping semantic structure, suggesting language-specific processes in organizing semantic lexicons. Since semantic subword components facilitate lexical-semantic processing (Ding et al., 2004; Feldman and Siok, 1999, and are more prevalent in Mandarin (Shen and Ke, 2007), it is likely that such components establish semantic connections between words in a more clustered manner.

With the network science approach, our work characterizes human semantic representation from a large-scale and global perspective. While few previous studies have considered different characteristics between different languages when studying lexical-semantic representation, our work underscores the impact of the diversity of language properties on human language processing. As studies have highlighted relations between semantic network properties and lexical properties (De Deyne et al., 2013; Peters and Borovsky, 2019) as well as semantic network structure and language processing Balota et al., 2004; Goñi et al., 2011; Mirman and Magnuson, 2008; Xu et al., 2021, our work may also provide insights into other aspects of human language processing, such as concepts, categorization and semantic memory of different populations. For example, as C of the lexicon has been shown to affect word production (Siew et al., 2019; Vitevitch and Luce, 2016; Xu et al., 2021) such that words with higher Cs might be harder to retrieve, there could be cross-linguistic differences in lexical retrieval due to different lexical-semantic structures.

Corpus analysis and cross-linguistic comparisons offer naturalistic and large-scale insights, but also introduce noise. While controlling for linguistic and other confounds, our study demonstrates correlational, not causal relationships between examined variables. Despite using text from comparable sources, content differences between languages could still exist and confound the connection between language and lexico-semantic structure. Future research could employ experimental methods to test language-specific properties' effects on language processing. Further development of cross-linguistic resources, such as psycholinguistic norms (Brysbaert et al., 2014) and word association norms (De Deyne et al., 2019; Nelson et al., 2004), would be valuable. Including more languages through such resources could provide better insights into the diversity of human language processing.

References

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological review*, *116*(3), 463.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *55*(2), 290–313.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of experimental psychology: General*, *133*(2), 283.

Bertail, P. (1997). Second-order properties of an extrapolated bootstrap without replacement under weak assumptions. *Bernoulli*, 149–179.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, *5*, 135–146.

Borge-Holthoefer, J., & Arenas, A. (2010). Semantic networks: Structure and dynamics. *Entropy*, *12*(5), 1264–1302.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental psychology*.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, *46*, 904–911.

Cai, Q., & Brysbaert, M. (2010). Subtlex-ch: Chinese word and character frequencies based on film subtitles. *PloS one*, *5*(6), e10729.

Cao, S., Lu, W., Zhou, J., & Li, X. (2018). Cw2vec: Learning chinese word embeddings with stroke n-gram information. *Proceedings of the AAAI conference on artificial intelligence*, *32*(1).

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior research methods*, *45*, 480–498.

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "small world of words" english word association norms for over 12,000 cue words. *Behavior research methods*, *51*, 987–1006.

Deng, Y., Wang, Y., Qiu, C., Hu, Z., Sun, W., Gong, Y., Zhao, X., He, W., & Cao, L. (2021). A chinese conceptual semantic feature dataset (ccfd). *Behavior Research Methods*, 1–13.

Devereux, B. J., Tyler, L. K., Geertzen, J., & Randall, B. (2014). The centre for speech, language and the brain (cslb) concept property norms. *Behavior research methods*, *46*, 1119–1127.

Ding, G., Peng, D., & Taft, M. (2004). The nature of the mental representation of radicals in chinese: A priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 530.

Feldman, L. B., & Siok, W. W. (1997). The role of component function in visual recognition of chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 776.

Feldman, L. B., & Siok, W. W. (1999). Semantic radicals contribute to the visual identification of chinese characters. *Journal of Memory and Language*, *40*(4), 559–576.

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage*, *32*(3), 1413–1421.

Garcıa, C. B., Salmerón, R., Garcıa, C., & Garcıa, J. (2020). Residualization: Justification, properties and application. *Journal of Applied Statistics*, *47*(11), 1990–2010.

Goñi, J., Arrondo, G., Sepulcre, J., Martincorena, I., Vélez de Mendizábal, N., Corominas-Murtra, B., Bejarano, B., Ardanza-Trevijano, S., Peraita, H., Wall, D. P., et al. (2011). The semantic organization of the animal category: Evidence from semantic verbal fluency and network theory. *Cognitive processing*, *12*, 183–196.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009a). Categorical structure among shared features in networks of early-learned nouns. *Cognition*, *112*(3), 381–396.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009b). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological science*, *20*(6), 729–739.

Hsiao, J., Shillcock, R., & Lavidor, M. (2007). An examination of semantic radical combinability effects with lateralized cues in chinese character recognition. *Perception & psychophysics*, *69*, 338–344.

Huang, C.-R., & Hsieh, S.-K. (2015). Chinese lexical semantics. In *The oxford handbook of chinese linguistics* (pp. 290–305). Oxford Univ. Press.

Kajic, I., & Eliasmith, C. (2018). *Evaluating the psychological plausibility of word2vec and glove distributional semantic models* (tech. rep.). Technical Report CTN-TR-20180824-012. Centre for Theoretical Neuroscience . . .

Karuza, E. A., Thompson-Schill, S. L., & Bassett, D. S. (2016). Local patterns to global architectures: Influences of network topology on human learning. *Trends in cognitive sciences*, *20*(8), 629–640.

Li & Kang, J. (1993). Analysis of phonetics of the ideophonetic characters in modern chinese. *Information analysis of usage of characters in modern Chinese*, 84–98.

Mair, P., & Wilcox, R. (2020). Robust statistical methods in r using the wrs2 package. *Behavior research methods*, *52*, 464–488.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Jour-*

nal of Experimental Psychology: Learning, Memory, and Cognition, *34*(1), 65.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

Onnela, J.-P., Saramäki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Physical Review E*, *71*(6), 065103.

Peters, R., & Borovsky, A. (2019). Modeling early lexico-semantic network development: Perceptual features matter most. *Journal of Experimental Psychology: General*, *148*(4), 763.

Shen, H. H., & Ke, C. (2007). Radical awareness and word acquisition among nonnative learners of chinese. *The Modern Language Journal*, *91*(1), 97–111.

Siew, C. S., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019). Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, *2019*.

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

Utsumi, A. (2015). A complex network approach to distributional semantic models. *PloS one*, *10*(8), e0136277.

Virpioja, S., Smit, P., Grönroos, S.-A., Kurimo, M., et al. (2013). Morfessor 2.0: Python implementation and extensions for morfessor baseline.

Vitevitch, M. S., Ercal, G., & Adagarla, B. (2011). Simulating retrieval from a highly clustered network: Implications for spoken word recognition. *Frontiers in psychology*, *2*, 369.

Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual review of linguistics*, *2*, 75–94.

Xu, Q., Markowska, M., Chodorow, M., & Li, P. (2021). A network science approach to bilingual code-switching. *Proceedings of the Society for Computation in Linguistics*, *4*(1), 18–27.

Yang, Z., Qi, P., Zhang, S., Bengio, Y., Cohen, W. W., Salakhutdinov, R., & Manning, C. D. (2018). Hotpotqa: A dataset for diverse, explainable multi-hop question answering. *arXiv preprint arXiv:1809.09600*.

Yeh, S.-L., Chou, W.-L., & Ho, P. (2017). Lexical processing of chinese sub-character components: Semantic activation of phonetic radicals as revealed by the stroop effect. *Scientific reports*, *7*(1), 1–12.

Yu, J., Jian, X., Xin, H., & Song, Y. (2017). Joint embeddings of chinese words, characters, and fine-grained subcharacter components. *Proceedings of the 2017 conference on empirical methods in natural language processing*, 286–291.