Composition as nonlinear combination in semantic space: Exploring the effect of compositionality on Chinese compound recognition

Tianqi Wang (tianqi93@connect.hku.hk)
Speech Science Laboratory, The University of Hong Kong
Pok Fu Lam, Hong Kong, China

Xu Xu (xu.xu@sjtu.edu.cn)
School of Foreign Languages, Shanghai Jiao Tong University
800 Dongchuan Rd., Shanghai, 200240, China

Abstract

Most Chinese words are compounds formed through the combination of meaningful characters. Yet, due to compositional complexity, it is poorly understood how this combinatorial process affects the access to the whole-word meaning. In the present study, we turned to the recent development in compositional distributional semantics (Marelli et al., 2017), and employed a deep neural network to learn the less-than-systematic relationship between the constituent characters and the compound words. Based on the compositional representations derived from the computational model, we quantified compositionality as the degree of overlap between the compositional and the lexicalized representations as well as the degree of distinctness of the compositional representation. We observed that these two compositional attributes can affect compound recognition over and above the effects of constituent character features and compound features. Moreover, we found that this effect was increasingly stronger when holistic access to the compound meaning became more challenging. These findings therefore, from a computational perspective, provided new evidence for the combinatorial process involved in Chinese word recognition, which also shed light on the universal process of compound comprehension.

Keywords: word recognition; compound word processing; Chinese word formation; compositionality; distributional semantic models

Introduction

Compound words, such as *swordfish*, allow speakers to create new expressions via the combination of existing elements. These words "are not just expressions that happen to embed other words but are structurally related to their constituents" (Günther & Marelli, 2019). In psycholinguistics, the semantic influences of constituent meanings on the processing of compound meaning are typically investigated through the measure of semantic transparency (ST). It has long been assumed that semantically transparent compounds (e.g., *swordfish*, as it refers to a type of *sword*-shaped *fish*) have processing advantages over those that are semantically opaque (e.g., *ladybird*, as it refers to a type of black-dotted bug that has little to do with the meaning of its constituents).

Empirical evidence for the ST effect was however inconsistent. In lexical decision studies, Libben et al. (2003) showed that compounds with transparent heads (i.e., the second constituent of the compound; e.g., *bedroom*, *strawberry*) were recognized faster than those whose head were opaque (e.g. *jailbird*, *hogwash*). Ji and colleagues (2011) observed that the ST effect would only emerge when a compositional process was encouraged (Experiment 4: inserting space between constituents, e.g., *rose bud*; Experiment 5: presenting constituents in different colors, e.g., *rose* in red and *bud* in black). These findings have motivated researchers to reconceptualize the measure of ST from a compositional perspective (Gagné & Spalding, 2009; Günther et al., 2019; Marelli et al., 2015). It is the compositionality of the compound, characterizing the extent to which the whole-word meaning faithfully reflects the semantic combination of the constituents, that determines whether the compound meaning can be accessed through its constituents.

Inspiring evidence for this view came from Marelli and Luzzatti (2012), who measured ST through ratings of compositionality (to what extent can a compound meaning be predicted from its constituents?) in addition to ratings of constituent relatedness (to what extent are the constituent meanings related to the compound meaning?). They observed that ST could moderate the effect of constituent frequency (reflecting how easily the meaning of a constituent can be accessed) in a lexical decision task only when this property was conceptualized from a compositional perspective.

With the development of distributional semantics, Marelli et al. (2017) proposed the CAOSS framework (Composition as Abstract Operation in Semantic Space), which allows an explicit quantification of the compositional ST. Their idea was to measure the similarity between the compositional and the lexicalized compound meanings, as well as those between the compositional and constituents' meanings within a distributional semantic space. To generate vector-based representations for the compositional meaning, they trained a regression model which learnt to optimally construct the observed compound meaning based on a linear combination of the constituents' meanings (both represented as vectors in a distributional semantic space). The compositional representation for the combination of any constituents could thus be generated based on the parameter matrices obtained through training. Using these meaning representations, they defined a set of measures that characterized ST from the compositional perspective. It was interesting to observe that these measures could explain a number of semantic effects in English compound processing.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

While the emerging picture from alphabetic languages (e.g., English in Marelli et al., 2017 and Günther & Marelli, 2019; German in Günther, Marelli, et al., 2020) has highlighted the compositional aspect of ST, research on how this property would influence Chinese word recognition is relatively impoverished. Chinese is a particularly interesting writing system to study this topic because compounding is the most common tool for Chinese word formation (Packard, 2000; over 70% of Chinese words are compounds; Institute of Language Teaching and Research, 1986; DeFrancis, 1989). Moreover, the constituent characters that serve as the building blocks for Chinese words have a number of unique features. Specifically, Chinese is a morpho-syllabic language, and there is usually a correspondence between a character, a syllable, and a morpheme. Unlike what is presented in alphabetic languages where morphemes have obscure boundaries (i.e., a morpheme can involve a variable number of letters and syllables), constituent characters are highly salient perceptual units (e.g., there are two characters in the compound 冰箱 *refrigerator*, each displayed in a box-shaped area, and pronounced as /bing1/ and /xiang1/, respectively). Because of that, morphological segmentation can be executed with minimal effort, which potentially allows rapid activation of the morphemic meaning (see Tsang & Chen, 2013a and 2013b for evidence from priming studies; see Tsang et al., 2018 and Tse et al., 2017 for evidence from megastudies of lexical decision). It has been assumed that after the activation of the constituent characters, a combinatorial process is actively involved to compute the compound meaning (i.e., the combinatorial route). On the other hand, constituent characters usually exhibit meaning ambiguity (Chen et al., 2023; e.g., 花 refers to both *flower* and *to spend*) and phonological inconsistency (Tan & Perfetti, 1999; e.g., 曾 is pronounced as /ceng2/ or /zeng1/) when appear in isolation. This made some researchers (e.g., Packard, 1999) to assume that Chinese words are recognized more efficiently through holistic processing (i.e., the holistic route).

In the present study, we intend to bridge this knowledge gap by exploring the effect of compositionality on Chinese compound processing. Specifically, we took into account two attributes associated with the end product of the combinatorial route. The first attribute is the extent to which the compositional meaning representation (i.e., semantic combination of the constituents) converges with the lexicalized representation (i.e., the whole-word meaning) of a compound word, which characterizes the essence of ST from the compositional perspective and potentially the relationship of the holistic route and the combinatorial route. The second attribute is the degree of accessibility of the compositional representation of a compound word, which characterizes the efficacy of the combinatorial route. These two attributes are expected to provide a new interpretative framework to understand the processing differences in word recognition. In terms of compound compositionality, when there is no direct access to the whole-word meaning and thus the holistic route does not automatically prevail, low compositionality would hinder compound processing as there is a competition between the combinatorial route and the holistic route. High compositionality would facilitate compound processing as the two routes were in sync. In terms of accessibility of the compositional meaning, the more distinct the end product of the combinatorial route, the easier it is to be accessed and activated during compound processing.

To quantify these two attributes, we turned to the recent development in compositional distributional semantics (Marelli et al., 2017) to obtain the vector representations for the compositional meaning. One complication to apply the CAOSS framework to Chinese morphology comes from the more complex and less-than-systematic relationship between the constituent characters. To illustrate, there are roughly five morphological structures in Chinese (Huang & Liao, 2017), which are coordinate (e.g., 城市 *city*, literally *city-city*), subordinate (e.g., 黑板 *blackboard*, literally *black-board*), verb-object (e.g., 扫地 *sweep the floor*, literally *sweep-floor*), verb-resultative (e.g., 说明 *illustrate*, literally *speak-clear*), and subject-predicate (e.g., 晚安 *good night*, literally *night-peace*). We therefore adopted the Notch model (Nonlinear Transformation of Chinese Embeddings; Tseng & Hsieh, 2022) to model the rather unsystematic relations between the constituent characters. Specifically, we employed a sophisticated deep neural network architecture (with millions of parameters to fine-tune; see below), which was trained to simulate the compositional process by nonlinearly combining the representations of the constituent characters within the semantic space. The model was able to capture the interplay between these constituents by learning their role-dependent meanings. Using the Notch-generated vectors, we quantified the compositionality and accessibility of the compound's compositional representation based on its distributional properties within the semantic space. We then examined how these attributes would influence lexical decision latencies using two-character words in MELD-SCH (Megastudy of Lexical Decision in Simplified Chinese; Tsang et al., 2018).

Method

Computational Model

Vector representation of the compositional compound meaning was derived via the Notch model (Tseng & Hsieh, 2022), which was trained to acquire the compounding rules (i.e., how constituent characters can be combined to generate the compound meaning). The model's architecture included a pre-trained MacBERT[1] (chinese-macbert-large; Cui et al.,

[1] While implementing the Notch model, we did not employ the pretrained BERT (bert-base-chinese) as in Tseng and Hsieh (2022), but opted to use MacBERT (chinese-macbert-large) that had more parameters, pre-trained with improved masked language model task, and could therefore represent language more accurately. Based on our pilot observation, composition metrics derived from the MacBERT-based Notch model could better predict lexical decision latencies. In addition, it is important to note that either BERT or MacBERT are character-based (i.e., inputs to the model are

2020) followed by a task-specific fully-connected layer. To compute the compositional meaning for a compound such as 冰箱 *refrigerator*, the model took the sequence of character constituents, 冰 *ice* and 箱 *box*, as input, and mapped the encoded final state corresponding to the [CLS] token (1024 dimensions) onto a vector of 300 dimensions. With a mean-squared-error loss function, the model was fine-tuned to generate a vector (i.e., the composition vector) that could approximate the compound's lexicalized embedding[2]. It is important to note that multiple encoder blocks within MacBERT kept mixing and warping the constituents' embeddings that were generated at the model's first layer. This process allowed the embeddings of the constituents to change according to their specific roles in the compound meaning (from a theoretical point of view, constituent characters modulated each other's meaning, which formed the word; Xu, 1994). Hence, the output of the [CLS] token at the final encoder block was by no means a simple linear combination of the constituent embeddings, but instead a more dynamic nonlinear combination of the role-dependent constituent embeddings.

We utilized a total of 500 thousand words to train and validate the Notch model, of which their observed embeddings (300 dimensions, pre-trained using skip-gram with negative sampling algorithm; Mikolov et al., 2013) were provided by Li et al. (2018)[3]. Specifically, we retrieved all words in the SUBTLEX-CH database (Cai & Brysbaert, 2010), with the remaining words selected sequentially from the embedding database. It is important to note that even though we focused our analysis on the two-character compounds in later sections, we did not restrict the length of the words for model training, as this allowed us to have a larger number of word sample, helping to avoid model overfit. More importantly, accommodating words of different lengths to train the compositional model may enhance its generalizability and facilitate the learning of the constituents' role-dependent meanings. Our dataset therefore included 11,043 words with one character, 212,268 words with two characters, 184,278 words with three characters, 65,581 words with four characters, and 26,830 words with five or more characters. The word sample was randomly split into a training set of 490 thousand words and a validation set of 10 thousand words. Special care was taken to match the proportion of words of different lengths within the two sets. All word embeddings were normalized into unit length before training and validation.

The model was trained for one epoch with a batch size of 8. AdamW optimizer (Loshchilov & Hutter, 2019) was used with a learning rate of 1e-4, betas of (0.9, 0.999), and a L2 weight decay of 0.01. The learning rate was first warmed up for 400 steps and then linearly decayed for the rest of the training.

Composition Metrics

Based on the compositional embeddings (i.e., the vectors generated by the computational model) and actual embeddings of the 500 thousand words in our dataset, two composition metrics are defined to characterize the end product of the combinatorial route. Specifically, we consider the distributional properties associated with the compositional embedding, and hypothesized that if a word's compound meaning is highly predictable given the combination of its constituents, the compositional embedding of the compound should be close to its actual embedding. Moreover, the distinctness of the compositional compound meaning should also be reflected in the distributional properties of its lexical neighborhood. In distributional semantics, the similarity between two words' meaning is computed as the cosine distance between their respective embeddings. Therefore, two computed metrics, the similarity between the compositional and the lexicalized meaning representations (SimCL) and the range of the similarities between the compositional representation and its lexical neighbors (SimRangeTop50), are defined as follows. A few compound examples with high versus low values on these metrics are shown in Table 1.

SimCL: the cosine distance between the compositional and the actual compound embeddings. It measures the extent to which a word's compositional meaning aligns with (or diverges from) its lexicalized meaning.

SimRangeTop50[4]: the range of the cosine distances of the top 0.01% neighbors (50 words among the 500 thousand words) that are closest to the compositional compound. This metric is used to capture the extent to which the compositional embedding has a distinct presence within the semantic space. That is, if the compositional representation is highly distinct, there should be relatively few lexicalized items, possibly only one, which are sufficiently close to it. In that case, lexical access should be swift. Conversely, if the closest and the 50th closest lexical neighbors of the compositional representation are of similar distances, the compositional meaning should be ambiguous, which can result in a delay in lexical access.

tokenized into single characters). Therefore, composition is always involved in the Notch architecture.

[2] In computational linguistics, *embeddings* are high-dimensional real-valued *vectors* that encode the words' semantic information. In this study, the word *embedding* and *vector* are used interchangeably.

[3] We did not follow the original implementation of the Notch model, which utilized the 100-dimensional word embeddings from Tencent AI lab (Song et al., 2018). We observed that many words in this dataset were coarse-grained as it was designed to be more task-oriented.

[4] In Tseng and Hsieh (2022), a similar computed metric, SimRange, was also derived from the similarity between the compositional compound and its 50 closest neighbors. While SimRange was defined as the difference of similarity between the 0.90 and 0.10 quantiles, SimRangeTop50 was defined as the difference between the max and min value. Based on our pilot analysis, SimRangeTop50 could better characterize the distributional properties of the compositional compound, and could better predict lexical decision latencies of real words in MELD-SCH.

Table 1: Compound examples with high and low values on the SimCL and SimRangeTop50.

Composition metrics	Low-value items (< the 10th percentile)	High-value items (> the 90th percentile)
SimCL	海绵 *sponge* (*sea* + *silk floss*)	乐器 *music instrument* (*music* + *equipment*)
	开方 *square root* (*open* + *square*)	寒冷 *cold* (*cold* + *cold*)
SimRangeTop50	红心 *red heart* (*red* + *heart*)	背后 *behind* (*back* + *back*)
	拂晓 *dawn* (*touch lightly* + *morning*)	复杂 *complex* (*repeat* + *miscellaneous*)

Analyses on Lexical Decision Task

To examine how lexical decision times were influenced by the computed metrics, we used the real word data from MELD-SCH (Tsang et al., 2018). This database comprises z-transformed response times (zRT; at the participant level) and error rates for a total of 12,578 Chinese words. Responses to these items were collected from 504 Chinese native speakers, and for each word, information on a series of psycholinguistic variables were also made available.

Here, we restricted our analyses to two-character compound words (10,022 in total), as the majority (about 73.6%) of modern Chinese words consist of two characters (Institute of Language Teaching and Research, 1986). In addition, we intended to add variables at the character level into our statistical model (see below). We first removed items whose error rates across participants was above 30%. After implementing this criterion, 9,627 words were left in total (drop rate = 3.94%). Further, because variables of interest may not be available for all words or their constituent characters, latencies for 9,620 real words (drop rate = 0.07%) eventually served as the dependent variable in the regression model.

In addition to the psycholinguistic variables provided in MELD-SCH, we computed family size (Hsieh et al., 2023; type-based and position-specific) for each constituent character using the SUBTLEX-CH frequency database (Cai & Brysbaert, 2010). Log-transformation were implemented to all variables of interest whenever appropriate. For variables serving as interaction terms, they were centered to mean before entering into the regression model.

Results

Before examining the efficacy of the composition metrics, we started from a baseline linear mixed effects model which included the set of lexical, semantic, and phonological variables that were known to predict word recognition. These variables were considered in the literature on compound processing (e.g., Günther & Marelli, 2019; Günther et al., 2020; Hsieh et al., 2023) as well as in the original megastudy (Tsang et al., 2018) based on which our analyses were conducted. We then included the composition metrics over and above the baseline variables, and examined whether they could significantly improve the model using likelihood-ratio tests. Overly influential outliers were first identified on the basis of a threshold of 2.5 units of standardized residual errors. The models were then refitted on the truncated dataset without outliers (model criticism; Baayen, 2008; see also Günther & Marelli, 2019).

For our analyses, linear mixed effects models were built with the *lme4* package (Bates et al., 2015). Subsequent analyses were implemented with the *MuMIn* (Bartoń, 2023) and *r2glmm* packages (Jeager, 2017) in R.

Baseline Model

Variables at the word and character level were taken into account as fixed effects of the baseline model. These variables included word frequency (LogWF) and number of strokes (NumStroke), as well as character frequency (LogCF), family size (LogFS), number of meanings (LogNoM), and number of pronunciations (LogNoP) of the first (C1) and second character (C2). Random intercepts for C1 and C2 were also included to capture the partial repeated-measure structure of the word sample, and to account for item variability without introducing an idiosyncratic effect for each item (since we had exactly one observation per item). The baseline model included all variables of interest (all ps < 0.010), except C1.LogNoM (p = 0.984) and C2.LogNoP (p = 0.057). Variance inflation factors for these variables ranged from 1.03 and 2.62 (as estimated by the R package *usdm*; Naimi et al., 2014), indicating that multicollinearity was not an issue for this model. The significance of variables at the character level was consistent with the results in previous megastudies (Tsang et al., 2018; Tse et al., 2017), suggesting that character features are activated during compound processing.

Efficacy of the Computed Metrics

To examine whether the computed metrics could explain unique variance in lexical decision latencies, we included the metrics over and above the baseline variables. As indicated by the likelihood ratio test, the inclusion of SimCL and SimRangeTop50 significantly improved the fit of the model, $\chi^2(2) = 233.59$, $p < 0.001$. Both metrics showed facilitatory effect on lexical decision latencies, as higher values were associated with faster responses. Moreover, there was a moderate correlation between the two metrics (r = 0.462), and the finding that SimRangeTop50 emerged as a significant predictor in addition to SimCL suggested that the two metrics captured at least some unique aspects of the compositional meaning, i.e., the end product of the combinatorial route.

Interaction with Other Predictors

Following previous studies (e.g., Günther & Marelli, 2019; Marelli & Luzzatti, 2012), we employed the same procedure (i.e., comparing one model with all the variables included so far to the other one with an extra predictor) to assess the potential interactions between the composition metrics (SimCL and SimRangeTop50) and the other variables (i.e., LogWF, as well as LogCF, LogFS, LogNoM, and LogNoP for C1 and C2). According to the likelihood ratio tests, the SimCL by LogWF, $\chi^2(1) = 21.23$, $p < 0.001$, SimRangeTop50 by LogWF, $\chi^2(1) = 19.11$, $p < 0.001$, SimRangeTop50 by C2.LogCF, $\chi^2(1) = 24.50$, $p < 0.001$, and SimRangeTop50 by C2.LogFS interactions, $\chi^2(1) = 4.55$, $p = 0.033$, explained additional variance in zRT, although the ones involving SimRangeTop50 and LogWF as well as SimRangeTop50 and C2.LogFS did not approach significance in the final model. Table 2 shows the added parameters and their respective contributions in terms of R^2 improvement over the previous predictors.

Table 2: Parameters of the final model.

Parameter	Estimate	*SE*	*t*	*df*	*p*	% ΔR^2	R^2
Intercept	-0.30	0.003	-90.82	943	< 0.001		
LogWF	-0.21	0.003	-68.17	9149	< 0.001		
Stroke	0.004	0.001	6.25	3044	< 0.001		
C1.LogCF	0.03	0.005	5.53	3028	< 0.001		
C2.LogCF	0.03	0.005	4.74	2220	< 0.001		
C1.LogFS	-0.06	0.009	-6.84	1788	< 0.001		
C2.LogFS	-0.06	0.010	-6.42	1318	< 0.001		
C2.LogNoM	0.06	0.014	4.24	1166	< 0.001		
C1.LogNoP	0.09	0.031	2.84	1323	0.005		
Baseline model							0.435
SimCL	-0.42	0.048	-8.65	9245	< 0.001	2.67	0.446
SimRangeTop50	-0.29	0.044	-6.72	8081	< 0.001	0.63	0.449
SimCL × LogWF	0.19	0.046	4.07	9111	< 0.001	0.27	0.451
SimRangeTop50 × C2.LogCF	0.19	0.043	4.45	7747	< 0.001	0.19	0.451
Composition metrics						3.80	0.451

Note: Degree of freedom (*df*) is estimated with the *lmerTest* package (Kuznetsova et al., 2017). From the baseline model, the contribution of each additional parameter over the previous predictors is displayed as ΔR^2 in percentage.

Discussion

Decades of psycholinguistic research has revealed the roles of constituent characters in compound processing (i.e., arguments for compound processing through a combinatorial route; e.g., Tsang & Chen, 2013b; Tsang et al., 2018; Tse et al., 2017). At the same time, there are also arguments that compounds should be processed as a whole (e.g., Bai et al., 2008; Packard, 1999). In the present study, we empirically tested these hypotheses by obtaining two composition metrics that characterized the distributional properties of the end product of the combinatorial route. Compositional meaning representation was derived from a computational model trained to acquire the compounding rules. Using a megastudy of word recognition (Tsang et al., 2018), we examined how these metrics, reflecting the ease of composition based on the constituent characters, would affect Chinese compound processing.

While establishing the baseline model, we demonstrate that characters do not simply act as orthographic codes to access compound meaning. Their lexical (i.e., character frequency), semantic (i.e., family size, number of meanings), and phonological features (i.e., number of pronunciations) are also shown to predict the response times for compound processing. Over and above the baseline variables, the main effects of the composition metrics provide strong evidence that after the activation of the constituent characters, a combinatorial process is involved to compute the meaning of the compound word. With respect to SimCL, processing is faster when the compositional meaning representation is closer to the word's lexicalized meaning representation. We interpret this as the benefits of compositionality: If constituent characters can be easily combined to indicate the whole-word meaning, the holistic route converges with the combinatorial route, resulting in an advantage in compound processing. This finding indicates that SimCL captures the essence of the compositional aspect of ST, i.e., the predictability of a compound meaning based on the meanings of the morphemes (Marelli & Baroni, 2015). It is also in line with the past research findings that in cases where the constituents can be easily integrated, conceptual combination is facilitated and is less cognitively demanding (Gagné & Spalding, 2009).

With respect to SimRangeTop50, a higher value implies that the compositional meaning representation as the end product of the combinatorial route has a distinct presence in the semantic space and thus can be easily activated in compound processing. In contrast, a lower value suggests that the compositional representation is in close proximity of many lexicalized neighbors, and to resolve the uncertainty brought by these distractors, processing cost can increase.

Our results also revealed that the computed metrics showed interactions with specific properties of the word as well as the constituent characters. Specifically, there is an interaction between SimCL and word frequency: For low-frequency words, the effect of SimCL is highly prominent, whereas for high-frequency words, this effect is weakened (see Figure 1a). This result is in line with the argument that holistic processing prevails for words with high frequency of usage (Tse, 2010), as they are more likely to be retrieved as a whole. (Cui et al., 2021; Shen et al., 2018). Because of that, whether the compositional and the lexicalized representations converge would not make much difference. Conversely, low-frequency words do not have such benefit, as there is a lack of readily retrievable holistic representation. Because their meaning must be analyzed in terms of morphemes, the extent to which the compositional representation deviates or aligns with the lexicalized representation would have a direct impact on the efficiency of compound processing.

The effect of SimRangeTop50, on the other hand, is qualified by an interaction with character frequency. That is, the facilitatory effect of SimRangeTop50 is much more prominent when the second constituent characters are low-frequency characters; when they are of high frequency, the facilitatory effect becomes dampened. This finding can be expected, as high-frequency second character is likely the constituent of one or more other high-frequency words (Tsang et al., 2018). Therefore, although the compositional embedding has a distinct presence in the neighborhood with less distractors, semantic activation of these high-frequency words could hinder the access to the target word (see also Amenta et al., 2015).

To summarize, our results shed further light on the processes involved in Chinese compound recognition. In particular, we observed that two attributes associated with the end product of the combinatorial route, i.e., compositionality and accessibility of the compositional representation, can influence the efficiency of compound processing. Our results indicate that the combinatorial route seems to be always in action when processing the compound word, although for high-frequency compounds, lexical representations are more readily accessible from the holistic route. For a writing system like Chinese whose morphemic information is extremely prominent, taking into account these two attributes can help resolve the debate on whether Chinese words are recognized holistically or through a combinatorial process implemented on the constituent characters. We believe that the compositional distributional model, as well as the derived computed metrics, can be applied to explain the processing differences among compound words in other languages, which can help understand compound representation more generally.

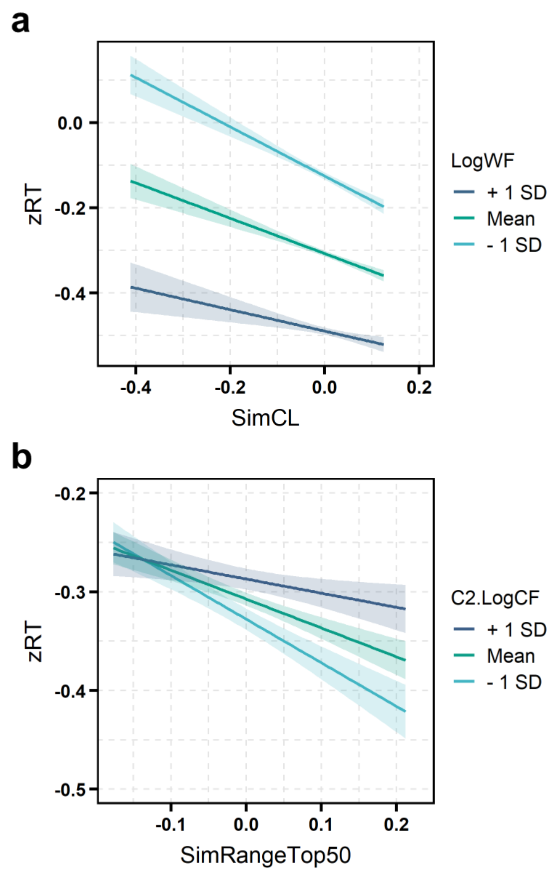

Figure 1: Interactions between (a) SimCL and LogWF and (b) SimRangeTop50 and C2.LogCF on lexical decision latencies. In each subplot, the moderator was discretized with representative values selected for visualization (mean and mean ± 1 SD according to Aiken et al., 1991). The moderator was treated as a continuous variable in the regression analysis.

References

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple Regression: Testing and Interpreting Interactions*: Sage.

Amenta, S., Marelli, M., & Crepaldi, D. (2015). The fruitless effort of growing a fruitless tree: Early morpho-orthographic and morpho-semantic effects in sentence reading. *Journal of Experimental Psychology: Learning Memory and Cognition, 41*(5), 1587-1596.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 34*(5), 1277-1287.

Bartoń, K. (2023). MuMIn: Multi-model inference (Version 1.47.5). Retrieved from https://cran.r-project.org/web/packages/MuMIn/index.html

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE, 5*(6), e10729.

Chen, H., Xu, X., & Wang, T. (2023). Assessing lexical ambiguity of simplified Chinese characters: Plurality and relatedness of character meanings. *Quarterly Journal of Experimental Psychology*.

Cui, L., Wang, J., Zhang, Y. L., Cong, F. J., Zhang, W. X., & Hyönä, J. (2021). Compound word frequency modifies the effect of character frequency in reading Chinese. *Quarterly Journal of Experimental Psychology, 74*(4), 610-633.

Cui, Y., Che, W., Liu, T., Qin, B., Wang, S., & Hu, G. (2020). *Revisiting pre-trained models for Chinese natural language processing.* Paper presented at the Findings of the Association for Computational Linguistics: Empirical Methods in Natural Language Processing 2020, Online.

DeFrancis, J. (1989). *Visible Speech: The Diverse Oneness of Writing Systems*. Hawaii, US: University of Hawaii Press.

Gagné, C. L., & Spalding, T. L. (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language, 60*(1), 20-35.

Günther, F., & Marelli, M. (2019). Enter Sandman: Compound Processing and Semantic Transparency in a Compositional Perspective. *Journal of Experimental Psychology: Learning Memory and Cognition, 45*(10), 1872-1882.

Günther, F., Marelli, M., & Bölte, J. (2020). Semantic transparency effects in German compounds: A large dataset and multiple-task investigation. *Behavior Research Methods, 52*(3), 1208-1224.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science, 14*(6), 1006-1033.

Hsieh, C. Y., Marelli, M., & Rastle, K. (2023). Beyond quantity of experience: Exploring the role of semantic consistency in Chinese character knowledge. *Journal of Experimental Psychology: Learning Memory and Cognition*.

Huang, B., & Liao, X. (2017). *Modern Chinese*. Beijing, China: Higher Education Press.

Institute of Language Teaching and Research. (1986) A frequency dictionary of Modern Chinese. Beijing: Beijing Language Institute Press.

Jeager, B. (2017). r2glmm: Computes R squared for mixed (multilevel) models (Version 0.1.2). Retrieved from https://github.com/bcjaeger/r2glmm

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language, 65*(4), 406-430.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26.

Li, S., Zhao, Z., Hu, R., Li, W., Liu, T., & Du, X. (2018). *Analogical reasoning on Chinese morphological and semantic relations.* Paper presented at the Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia.

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language, 84*(1), 50-64.

Loshchilov, I., & Hutter, F. (2019). *Decoupled weight decay regularization.* Paper presented at the Proceedings of the 7th International Conference on Learning Representations, New Orleans, US.

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review, 122*(3), 485-515.

Marelli, M., Dinu, G., Zamparelli, R., & Baroni, M. (2015). Picking buttercups and eating butter cups: Spelling alternations, semantic relatedness, and their consequences for compound processing. *Applied Psycholinguistics, 36*(6), 1421-1439.

Marelli, M., Gagné, C. L., & Spalding, T. L. (2017). Compounding as Abstract Operation in Semantic Space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition, 166*, 207-224.

Marelli, M., & Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory and Language, 66*(4), 644-664.

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

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography, 37*(2), 191-203.

Packard, J. L. (1999). Lexical access in Chinese speech comprehension and production. *Brain and Language, 68*, 89-94.

Packard, J. L. (2000). *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge, UK: Cambridge University Press.

Shen, W., Li, X. S., & Pollatsek, A. (2018). The processing of Chinese compound words with ambiguous morphemes in sentence context. *Quarterly Journal of Experimental Psychology*, *71*(1), 131-139.

Song, Y., Shi, S., Li, J., & Zhang, H. (2018). *Directional skip-gram: Explicitly distinguishing left and right context for word embeddings.* Paper presented at the Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, US.

Tan, L., & Perfetti, C. A. (1999). Phonological activation in visual identification of Chinese two-character words. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*, 382-393.

Tsang, Y. K., & Chen, H. C. (2013a). Early morphological processing is sensitive to morphemic meanings: Evidence from processing ambiguous morphemes. *Journal of Memory and Language, 68*, 223-239.

Tsang, Y. K., & Chen, H. C. (2013b). Morpho-semantic processing in word recognition: Evidence from balanced and biased ambiguous morphemes. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(6), 1990-2001.

Tsang, Y. K., Huang, J., Lui, M., Xue, M. F., Chan, Y. W. F., Wang, S. P., & Chen, H. C. (2018). MELD-SCH: A megastudy of lexical decision in simplified Chinese. *Behavior Research Methods*, *50*(5), 1763-1777.

Tse, C. S. (2010). *Interactive effect of semantic transparency and word frequency on lexical decision for Chinese compound words.* Paper presented at the Proceedings of the 51st Annual Meeting of the Psychonomic Society, St. Louis, US.

Tse, C. S., Yap, M. J., Chan, Y. L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods, 49*(4), 1503-1519.

Tseng, Y. H., & Hsieh, S. K. (2022). *Character Jacobian: Modeling Chinese character meanings with deep learning model.* Paper presented at the Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Korea.

Xu, T. (1994). Chinese character and research methodology: A commentary on the westernized view-points in Chinese linguistic studies. *Chinese Teaching in the World, 29*(3), 1-14.