Making "concreteness" more concrete

Yanting Li, Gregory Scontras, and Richard Futrell
Department of Language Science
University of California, Irvine
{yantil5, g.scontras, rfutrell}@uci.edu

1 Introduction

Concrete words (e.g., *apple* as opposed to *appetite*) are often described as more specific (Paivio, 1963) and perceptible (Lambert, 1955), with a more direct reference to objects, material, or sources of sensation (Gorman, 1961; Paivio, 1966). Concreteness is an important concept in the field of psycholinguistics. Previous research has found that concrete words are generally more easily learned (Paivio, 1963), more easily retrieved for lexical decision tasks (Schwanenflugel and Akin, 1994), and they tend to elicit more associated words (Lambert, 1955) with shorter response time (Richardson, 1975; Paivio, 1966). Different theories have been proposed to explain this concreteness effect. The Dual Coding Theory proposes that concrete words have nonverbal (perceptual) representations in addition to verbal ones, facilitating processing (Paivio, 1986). The Context Availability Theory, on the other hand, proposes that concrete words have more relevant contexts in people's prior knowledge that aid processing (Schwanenflugel et al., 1988).

In bilingual populations, concrete words have been shown to elicit faster lexical decision response times in both languages (van Hell and de Groot, 1998a) and elicit more similar lexical associations across languages than abstract words (Taylor, 1976; van Hell and de Groot, 1998b). Additionally, they have been found to facilitate code-switching (Myslín and Levy, 2015). One explanation is that concrete words share more semantic features across languages, resulting in more similar cross-linguistic associations and easier co-activation of translation equivalents (van Hell and de Groot, 1998b; Kroll et al., 2006). Abstract words, by contrast, are more dependent on the linguistic context in which they appear rather than on perceptual properties (Breedin et al., 1994), making them more likely to have language-specific meanings and trigger different associations across languages (van Hell and de Groot, 1998b).

Are concrete words more similar across languages than abstract words? Multilingual aligned word vectors offer a useful way to test this question, as they locate words from different languages in a shared embedding space by learning the contexts in which they typically occur (Mikolov et al., 2013; Smith et al., 2017; Conneau et al., 2017). In this shared space, distance between any two vectors can indicate their semantic relatedness. If concrete words do share more semantic features across languages, they should be located closer to their nearest neighbor in another language than abstract words. This is the hypothesis we aim to test.

2 Method

Concreteness ratings We found previously-collected concreteness ratings of words in six languages: Dutch (Brysbaert et al., 2014a), English (Brysbaert et al., 2014b), French (Bonin et al., 2018), Cypriot Greek (Traikapi et al., 2024), Mandarin (Xu and Li, 2020) and Portuguese (Soares et al., 2017) (Table 1). These languages were chosen because they are among the 44 languages that have pre-trained aligned word vectors (Joulin et al., 2018; Bojanowski et al., 2017). For all six datasets, concreteness was rated by self-identified adult native speakers recruited from universities or through crowd-sourcing platforms. "Concrete word" were generally described as words that can be experienced, perceived or learned through the senses. Each word was typically shown in isolation without linguistic context[1] and received around 30 rat-

[1]This could introduce the problem of "concreteness ambiguity" for lexical items with more than one meaning that have different degrees of concreteness (Löhr, 2021). For example, *table* referring to a piece of furniture is relatively concrete, whereas *table* referring to a set of values in a document may be less so. However, pretrained word embeddings also bear such ambiguity, as the embedding of a word form reflects the aggregate of all contexts it appears in. We therefore do not expect such ambiguity to systematically bias the relationship

Proceedings of the Society for Computation in Linguistics (SCiL) 2026, pages 338–342

Language	Total word count	Words with vector	Word information
Dutch	30,070	28,095	lemmas from all syntactic categories
English	39,954	35,318	lemmas from all syntactic categories
French	1,659	1,576	67% nouns, plus verbs, adjectives, adverbs and function words
Cypriot Greek	2,948	327	mostly nouns, with a few adjectives and verbs
Mandarin	9,877	8,421	two-character simplified Chinese words, excluding prepositions
Portuguese	3,800	3,770	content words including nouns, adjectives and others

Table 1: Concreteness rating information for the six languages used in the study.

ings, with counts ranging from 15 to 62 across languages.

Vector distance For each word with a rating, we located it in the 44-language embedding space with pre-trained multilingual aligned word vectors (Joulin et al., 2018; Bojanowski et al., 2017) and measured both the Euclidean distance and cosine similarity to its nearest neighbor among words in the other 43 languages.[2] We focused on nearest neighbors rather than translation equivalents because the latter is often unavailable or unstable for abstract words (van Hell and de Groot, 1998b). Even given a concrete word, there might be multiple translation equivalents. Therefore, identifying and determining translation equivalents for each word across 44 languages would require substantial language-specific judgment and introduce additional subjectivity into the analysis. Using the nearest neighbor instead could provide a consistent and fully data-driven approximation of cross-linguistic semantic similarity. As Euclidean distance and cosine similarity between two vectors were strongly correlated, analyses based on either metric yielded similar results. In the remainder of the paper, we report results based on Euclidean distance.

Word frequency Word frequency may influence both concreteness ratings and vector distances. For example, function words tend to be both highly frequent and relatively abstract; more frequent words generally have more stable embedding representations. We therefore included word frequency in our analyses. The English concreteness rating dataset included frequency information from SUBTLEX-US (Brysbaert and New, 2009). To maintain consistency across languages, we added frequency information for Dutch and Mandarin using SUBTLEX-NL (Keuleers et al., 2010) and SUBTLEX-CH (Cai and Brysbaert, 2010), respectively. For the remaining three languages for which SUBTLEX-based frequency norms were not available, log frequency estimates were obtained using the `wordfreq` Python library (Speer, 2022).

Analysis For each language with concreteness ratings, we fit the following linear regression model 43 times, each time using a different language as the source of nearest-neighbor distances:

`concreteness` $\sim$ `distance + word frequency`

3 Results

Fig 1 illustrates the relationship between word concreteness ratings and the distance to the nearest neighbor across 44 languages. The six panels represent the six languages for which concreteness ratings are available, and each line within a panel corresponds to one of the 43 neighboring languages. Solid lines indicate that distance was a significant predictor in the regression model described above ($p < .05$), whereas dashed lines indicate non-significant effects. The relatively lower proportion of significant effects observed for Cypriot Greek (91%) and French (44%) may partly reflect smaller effective sample sizes, as these datasets are relatively limited to begin with and the analyses additionally require the availability of pretrained aligned embeddings for each word.

The relationship between concreteness rating and distance varies considerably across languages. For English and Mandarin, the correlations are uniformly negative and significant across all neighboring languages. For French, the correlation is significant for 19 of the 43 neighboring languages, and all significant effects are negative. In contrast, Dutch, Cypriot Greek, and Portuguese generally exhibit positive correlations, with a few exceptions (Chinese and Vietnamese for Portuguese; and additionally Arabic for Dutch).

between concreteness and cross-linguistic semantic distance.

[2]Because aligned word vectors are trained on large multilingual corpora, the vocabulary of a given language may contain words from other languages. To minimize such cross-language contamination, we restricted searches to words written in the dominant script of the target language. For example, when Bulgarian was the neighboring language, the nearest neighbor must be written in Cyrillic script.

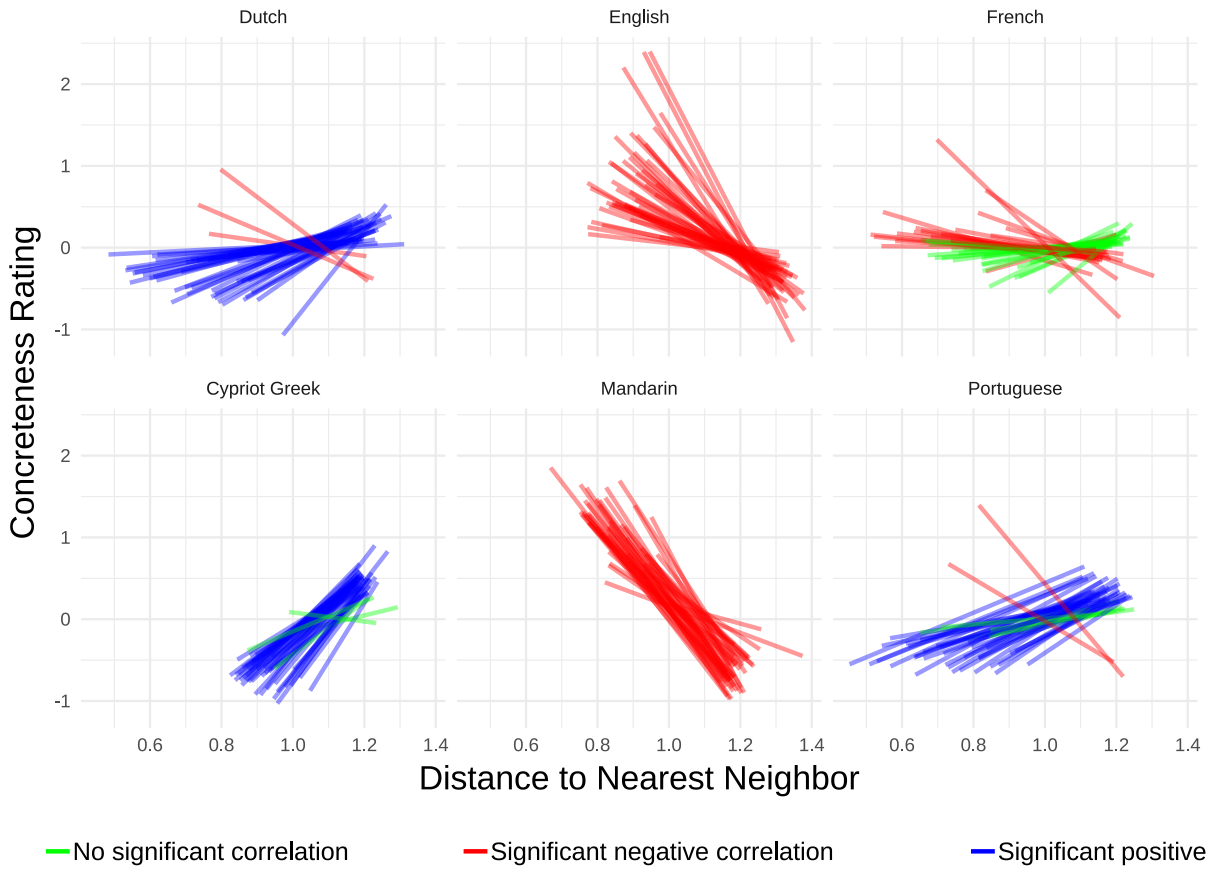

Figure 1: Relationship between normalized concreteness ratings (y axis) and distance to nearest cross-linguistic neighbor (x axis) across six rating languages (panels). Lines show simple linear regressions (concreteness ∼ distance), with color indicating the direction and significance of the distance effect in the multiple regression controlling for word frequency.

4 Discussion

Our results suggest a lack of a uniform relationship between cross-linguistic semantic relatedness and concreteness ratings across languages. English, Mandarin, and part of the French data are consistent with the hypothesis that concrete words are more similar cross-linguistically. However, Dutch, Portuguese, and Cypriot Greek showed the opposite pattern, suggesting that concrete words may be more language-specific than less concrete words.

One possible interpretation is that speakers of different languages associate concreteness with different semantic properties, despite the similar instructions used to elicit the ratings across languages. To examine this possibility more closely, we could identify overlapping concepts from different ratings datasets and check if their ratings are positively correlated. Another possible source of variation is the lexical composition of the ratings datasets, as some cover a broader range of syntactic categories, including highly abstract function words, whereas some do not, resulting in discrepancies among the effective endpoints of the rating scale. However, this explanation cannot fully account for the contrasting patterns of English and Dutch, whose datasets are comparatively similar in size and lexical coverage. Further investigation is therefore needed to determine which properties of the datasets or languages give rise to these differences. More broadly, the present findings highlight the importance of examining what "concreteness" captures for speakers of different languages.

While prior work has demonstrated that distributional semantic representations can be used to predict concreteness ratings (Thompson and Lupyan, 2018; Charbonnier and Wartena, 2020), the present study takes a more explanatory approach. Rather than treating embeddings as a predictive tool based on existing concreteness norms, we use them to probe what concreteness captures across languages and to what extent it reflects shared semantic or-

ganization. Through this attempt, we demonstrate how word embeddings can be used to operationalize psycholinguistic constructs involving cross-linguistic semantic relatedness, providing a quantitative complement to traditional ratings-based measures.

References

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.

Patrick Bonin, Alain Méot, and Aurélia Bugaiska. 2018. Concreteness norms for 1,659 French words: Relationships with other psycholinguistic variables and word recognition times. *Behavior Research Methods*, 50(6):2366–2387.

Sarah D Breedin, Eleanor M Saffran, and H Branch Coslett. 1994. Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology*, 11(6):617–660.

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.

Marc Brysbaert, Michaël Stevens, Simon De Deyne, Wouter Voorspoels, and Gert Storms. 2014a. Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, 150:80–84.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014b. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3):904–911.

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

Jean Charbonnier and Christian Wartena. 2020. Predicting the concreteness of german words. In *SWISSTEXT & KONVENS 2020: Swiss Text Analytics Conference & Conference on Natural Language Processing 2020; Proceedings of the 5th Swiss Text Analytics Conference (SwissText) & 16th Conference on Natural Language Processing (KONVENS), CEUR Workshop Proceedings Vol. 2624*.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2017. Word translation without parallel data. *arXiv preprint arXiv:1710.04087*.

Aloysia M Gorman. 1961. Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, 61(1):23.

Armand Joulin, Piotr Bojanowski, Tomáš Mikolov, Hervé Jégou, and Edouard Grave. 2018. Loss in translation: Learning bilingual word mapping with a retrieval criterion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2979–2984.

Emmanuel Keuleers, Marc Brysbaert, and Boris New. 2010. SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3):643–650.

Judith F Kroll, Susan C Bobb, and Zofia Wodniecka. 2006. Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9(2):119–135.

Wallace E Lambert. 1955. Associational fluency as a function of stimulus abstractness. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 9(2):103.

Guido Löhr. 2021. What are abstract concepts? on lexical ambiguity and concreteness ratings. *Review of Philosophy and Psychology*, 13:549—-566.

Tomas Mikolov, Quoc V Le, and Ilya Sutskever. 2013. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*.

Mark Myslín and Roger Levy. 2015. Code-switching and predictability of meaning in discourse. *Language*, 91(4):871–905.

Allan Paivio. 1963. Learning of adjective-noun paired associates as a function of adjective-noun word order and noun abstractness. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 17(4):370.

Allan Paivio. 1966. Latency of verbal associations and imagery to noun stimuli as a function of abstractness and generality. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 20(4):378.

Allan Paivio. 1986. *Mental Representations: A Dual Coding Approach*, chapter 4. Oxford University Press, New York.

John TE Richardson. 1975. Concreteness and imageability. *The Quarterly Journal of Experimental Psychology*, 27(2):235–249.

Paula J Schwanenflugel and Carolyn E Akin. 1994. Developmental trends in lexical decisions for abstract and concrete words. *Reading Research Quarterly*, (3):251–264.

Paula J Schwanenflugel, Katherine Kip Harnishfeger, and Randall W Stowe. 1988. Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, 27(5):499–520.

Samuel L. Smith, David H. P. Turban, Steven Hamblin, and Nils Y. Hammerla. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax.

Ana Paula Soares, Ana Santos Costa, João Machado, Montserrat Comesaña, and Helena Mendes Oliveira. 2017. The Minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods*, 49(3):1065–1081.

Robyn Speer. 2022. rspeer/wordfreq: v3.0.

Insup Taylor. 1976. Similarity between French and English words—a factor to be considered in bilingual language behavior? *Journal of Psycholinguistic Research*, 5(1):85–94.

Bill Thompson and Gary Lupyan. 2018. Automatic estimation of lexical concreteness in 77 languages. In *Proceedings of the annual meeting of the cognitive science society*, volume 40, pages 1120–1125.

Artemis Traikapi, Michaela Michael, and Nikos Konstantinou. 2024. The Greek-Cypriot word pool: Normative data for concreteness and imageability for 2,950 words.

Janet van Hell and Annette MB de Groot. 1998a. Disentangling context availability and concreteness in lexical decision and word translation. *The Quarterly Journal of Experimental Psychology: Section A*, 51(1):41–63.

Janet G van Hell and Annette MB de Groot. 1998b. Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1(3):193–211.

Xu Xu and Jiayin Li. 2020. Concreteness/abstractness ratings for two-character Chinese words in MELD-SCH. *PloS One*, 15(6):e0232133.