Title

Conceptual Diversity Across Languages and Cultures: A Study on Common Word Meanings among native English and Chinese speakers.

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Ke, Jia
De Deyne, Simon

Publication Date

2023-12-31

Peer reviewed

Conceptual Diversity Across Languages and Cultures: A Study on Common Word Meanings among native English and Chinese speakers.

Jia Ke (jkke1@student.unimelb.edu.au) and Simon De Deyne (simon.dedeyne@unimelb.edu.au)
School of Psychological Sciences, University of Melbourne, Australia

Abstract

While meaning variation in common words across language and culture is well established, only a few studies have explicitly quantified how general such differences are and whether differences reflect slight variations in meaning or could be considered to map onto entirely distinct concepts for different groups. The present study aims to investigate the extent to which common words can be interpreted differently between groups of English-proficient native Chinese speakers and native English speakers. This was done through a free judgment of associative strength (JAS) task using 42 cue English nouns. Our findings revealed language-specific meanings across all 42 cue words, with strong evidence for language-specific meaning in nearly 95% of nouns. To determine whether these words map onto entirely distinct language-specific concepts, we measured conceptual diversity using Latent Profile Analysis (LPA). The results of the LPA showed that nearly 69% of the cue words could be mapped onto more than one concept across all participants. Importantly, language differences were related to conceptual diversity in nearly 64% of words featuring multiple concepts. In sum, we found robust evidence of word meanings and conceptual variations among individuals across distinct linguistic and cultural backgrounds, even for common English words.

Keywords: bilingualism; culture; word meanings; conceptual diversity

Introduction

Communication relies on languages built upon a shared understanding of the words' meanings. However, recent work has shown that even common terms can exhibit systematic meaning variations across languages (Thompson, Roberts, & Lupyan, 2020; Vivas, Montefinese, Bolognesi, & Vivas, 2020). Understanding how common words vary in meaning across languages and cultures holds significant practical importance, as it can result in translational discrepancies and misinterpretations. For example, in English, the term *impressed* commonly has a positive valence. In contrast, its Italian translational-equivalent word, *impressionato*, has a more negative emotional valence, which might result in a semantic conflict when English is learned as a second language (Fairfield, Ambrosini, Mammarella, & Montefinese, 2017). It is also important theoretically, as models of bilingualism, such as the Revised Hierarchical Model (Kroll, Van Hell, Tokowicz, & Green, 2010), assume shared conceptual representations. whereas the more recent Modified Hierarchical Model leaves room for language-specific concepts (Pavlenko, 2007).

Recent work suggests that variation in concepts is widespread, even among monolinguals. For example, Martí, Wu, Piantadosi, and Kidd (2023) showed that words referring to names for animals or politicians mapped on several distinct concepts. One original aspect of this work is that multiple concepts mapping onto a single word were identified by clustering representational vectors derived from an adjective-based *judgment of associative strength* (JAS) task. This was based on a large group of participants using methods that group together response profiles for individuals to identify clusters of similar participants and automatically determine the number of clusters. Their results showed that a single cue word within the domain of animals can be associated with 5 to 8 distinct concepts. While this provides initial evidence for conceptual diversity for common words, the study did not address what lexical or demographic factors contribute to this conceptual diversity across individuals.

One potential source is gender, as previous studies have shown that men and women often have distinct conceptual representations of tool or fruit terms, resulting in differences in naming speed and preferences across genders (e.g., Capitani, Laiacona, & Barbarotto, 1999). Using a similar method to Martí et al. (2023), De Deyne, Warner, and Perfors (2023) investigated to what degree gender might predict conceptual diversity for a set of common English nouns. They found evidence for gender differences in about 30% of words that varied in concreteness, emotional valence, and gender stereotypically. Moreover, when determining the number of distinct concepts, they found that the conceptual diversity could be explained in terms of gender differences in 46% of these concepts. Altogether, this suggests that deriving individual representations using JAS and identifying concepts using clustering provides a measurement of conceptual diversity sensitive to gender differences. However, while gender is an important explanatory factor of meaning difference and conceptual diversity, it remains unclear to what degree other socio-demographic factors play a role. In the case of variation due to language background, few studies have directly addressed the question of whether this results in a continuum of meaning differences or whether the meaning of some words among bilinguals with a certain level of proficiency is sufficiently different to suggest distinct concepts.

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.*

Current study The current study uses a JAS task similar to Martí et al. (2023) and De Deyne et al. (2023) to obtain representations sensitive to patterned variation in meaning and concepts as a function of language background. In contrast to Martí et al. (2023), the current work aims to investigate a variety of words and concepts, primarily through varying word concreteness and valence, as these have been identified as primary factors that could drive meaning variation. For example, Wang and Bi (2021) found that concrete words referring to tangible objects (e.g., *cat*) exhibited less semantic variation than abstract words, as the latter lacks a concrete external reference (e.g., *business*). Furthermore, there is a growing recognition that emotion valence can exhibit systematic variations in meaning across languages. For example, a cross-cultural study found that the greater the geographic distance between languages, the less similarity existed in the emotional concepts associated with the same words (Jackson et al., 2019). Similarly, words with higher emotional valence were less semantically aligned across languages and cultures compared to neutral emotion words (Plutchik, 2001; Jackson et al., 2019).

The present study investigates non-native English speakers' understanding of common English word meanings by conceptually replicating and extending previous work on conceptual diversity and gender by De Deyne et al. (2023). The study uses the same set of 42 common English nouns as stimuli but changes the constrained adjective JAS used in De Deyne et al. (2023) to a free JAS where all part-of-speech are permitted. This study first aims to quantify the number of cue words that can be interpreted differently between native Chinese and English speakers. The second aim is to quantify the number of distinct concepts mapped onto the cue words using the latent profile analysis (cf. Martí et al., 2023). Beyond these two central aims, we will also consider the role of gender and lexical covariates (i.e., concreteness and valence) on semantic differences across all participants.

Method

Participants. A total of 187 participants (104 females and 83 males) aged between 17 and 35 (M_{age} = 19.6) completed the task. They were all first-year undergraduate students at the University of Melbourne who received course credits for their participation. Participants were screened based on their language and cultural background using the LEAP-Q. Only Chinese participants whose first and dominant language was Chinese, who culturally identified as Chinese, and who resided for less than 7 years in Australia were retained. The English speakers were screened to have English as their first and dominant language and have resided in Australia for over 10 years. Other inclusion criteria were based on the participants' Lextale vocabulary score and reliability across familiarity and JAS ratings (see below). The final sample comprised 152 participants (46 and 37 English female/male and 38 and 31 Chinese female/male participants) with the same age descriptives as before.

Materials and Measures. The cue words used in this study comprised a set of 42 words previously used in De Deyne et al. (2023) [1]. Cues were drawn from the Glasgow Norms dataset, and the selection criteria were based on normed ratings for gender (i.e., the degree to which the words may be masculine or feminine on a 7-point scale), concreteness (7-point abstract-concrete scale), and valence (9-point negative-positive scale) (Scott, Keitel, Becirspahic, Yao, & Sereno, 2019). Across the three levels of gender (feminine, masculine, neutral), independent t-tests confirmed that concreteness and valence were balanced (De Deyne et al., 2023).

The associated words for each cue word were extracted from the responses from the English Small World of Words (SWOW) word association norms (De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2019). Each cue word had at least 14 to 50 different associates. The associates were further manually screened for words that could be considered offensive, jargon, or were too similar to the cue word (e.g., *glory* and *glorious*) or orthographically similar to other responses. In cases where two orthographically related forms were available, the form with the largest association frequency was retained.

Procedure. Participants completed four tasks online in a 90-minute study. The first task involved a familiarity judgment task, where participants provided familiarity ratings for each cue word. Familiarity ratings were assessed using a 9-point scale, with 1 indicating "very unfamiliar" and 9 indicating "very familiar". The order of the cue words was randomised. Secondly, participants conducted the JAS task, evaluating the degree of association between a set of associates and each cue word. The JAS task was based on a continuous rating scale from 0 = no association to 100 = very strong association. Participants were instructed to use their intuition to evaluate the degree of association between a set of associates and each cue word, using the rating scale in an absolute sense.

The rating scale design presented all these associated words simultaneously, allowing participants to arrange and position them relative to one another. They were asked to begin with the associated word they believed was most strongly associated with the cue word. Additionally, the scale enabled zooming and panning using the mouse to locate words more accurately. The order of associates in each cue word trial was also randomized. At the end of the study, we administered a standardized Lextale vocabulary test (Lemhöfer & Broersma, 2012) to assess participants' English proficiency, along with a shortened version of the LEAP-Q questionnaire to examine their English language experience and cultural background (Marian, Blumenfeld, & Kaushanskaya, 2007).

Results

Data processing and checks were applied for three tasks within the experiment: familiarity ratings, the judgment of as-

[1] All data are available at `https://osf.io/bk2zw/`.

sociated strength (JAS) task, and the Lextale test. Participants with a normalized Lextale score (see Lemhöfer & Broersma, 2012) below 0.5 were excluded from further analysis, and participants who correlated less than .10 with the mean familiarity ratings or mean JAS ratings were also excluded from the analysis. In total, thirty-five participants were excluded from all further analyses. Next, we removed seven associations (*clandestine*, *opulent*, *echidna*, *Freud*, *hubris*, *marsupial*) that were unknown by more than 25% of Chinese speakers.

The familiarity ratings were significantly different between Chinese (M = 6.12, SD = 2.09) and English speakers (M = 7.45, SD = 0.93), $t(113.7) = 3.79$, $p < .001$. The familiarity ratings were highly reliable according to Spearman-Brown split-half reliability: r_{sb} = .94 and .89 for female/male English speakers and r_{sb} = .96 and .93 for female/male Chinese speakers. The JAS ratings were also highly reliable, r_{sb} = .90 and .86 for female/male English speakers, and r_{sb} = .87 and .81 for female/male Chinese speakers.

Quantifying cross-lingual alignment

Cross-lingual alignments were calculated to determine to what degree the meaning of a word was similar across native English speakers and non-native English speakers. Before the analysis, missing values for unknown associated words were replaced with the average judgment across all participants. Participants' ratings and cue words in the JAS task were standardized using z-scores. For each cue word, we calculated the alignment between the response ratings across all participants. To do so, we calculated all pairwise Euclidean distances within each group of English and Chinese participants. Next, we calculated the semantic alignment score for a specific participant. This score is determined by the correlation of an individual's semantic distance vector with the centroid vector, which averages distance scores for all other English or Chinese speakers. A response was classified as correct if the semantic alignment score was larger in the participant's language compared to the alternative language.

To quantify the degree of semantic alignment at the cue word level, we applied a Bayesian proportion test that compares a model where the number of successes is larger than chance with a null interval in which the number of successes equals or is less than chance. The results, with BF interpretations are shown in Figure 1. All 42 cue words had $BF_{10} > 1$, suggesting at least anecdotal evidence for language-specific meanings; 40 words (95.2%) with at least strong evidence, and 35 words (83.3%) had $BF_{10} > 30$, suggesting at least very strong evidence, of which 31 words (73.8%) exhibited extreme evidence, with $BF_{10} > 100$.

An additional analysis was conducted to investigate the role of gender by assessing how closely individuals' response ratings aligned with the average ratings of both males and females. As shown in Figure 2, a total of 8 words (19.0%) had $BF_{10} > 1$, suggesting at least anecdotal evidence; 3 words (7%) had $BF_{10} > 3$, suggesting at least moderate evidence. Next, to determine whether meaning variation was shared across gender and language, we compared the evidence strength for all 42 words. No significant correlation was found between the BF_{10} for language-specific meanings and gender-specific meanings, $r(40) = .06$, $p = .690$.

Determinants of cross-linguistic alignment We also explored whether semantic distance was related to word concreteness and valence differences among cue words. The analysis was performed by conducting a Welch t-test between the logBF for language-specific meanings of concrete/abstract and neutral/emotional words. The comparison between concrete words (M = 8.92) and abstract words (M = 10.64) was not significant, $t(37.21) = -0.88$, $p = .386$. Similarly, affective words (M = 9.36) and neutral words (M = 10.33) were also not significantly different, $t(30.79) = -0.47$, $p = .642$.

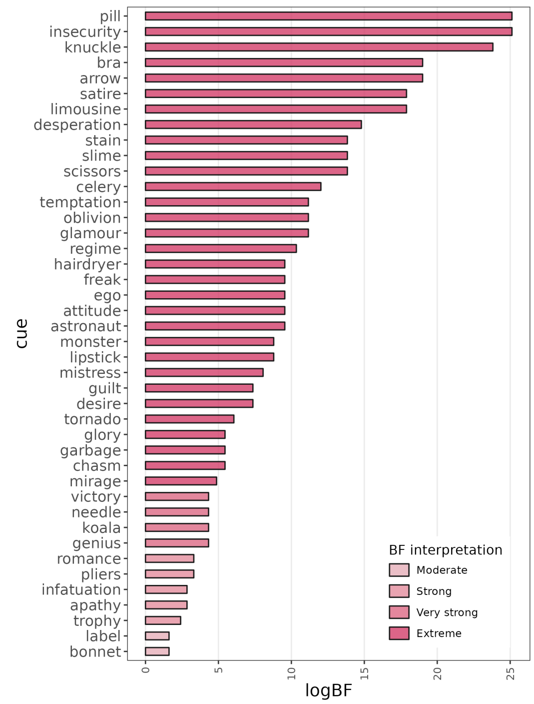

Figure 1: Evidence for language-specific meaning with BF interpretations according to Jeffreys (1961).

One potential alternative explanation for the observed meaning differences is that participants might not be familiar with the meanings of the words. To explore this possibility, we calculated the difference between the average familiarity ratings of the English and Chinese participant groups for each cue word. Then, we compared this difference with the logBF for language-specific meanings. The correlation was non-significant, $r(40) = .09, p = .566$. Next, we calculated whether an individual participant's familiarity ratings were correlated with the semantic alignment score toward the English and Chinese centroids. We only found a weak correlation between individual familiarity ratings and the English centroid, $r(6400) = .15, p < .001$, while the correlation with the Chinese centroid was non-significant, $r(6400) = .01, p = .346$. This suggests that word familiarity only plays a minor role in explaining meaning differences.

Finally, we assessed whether word meaning differences

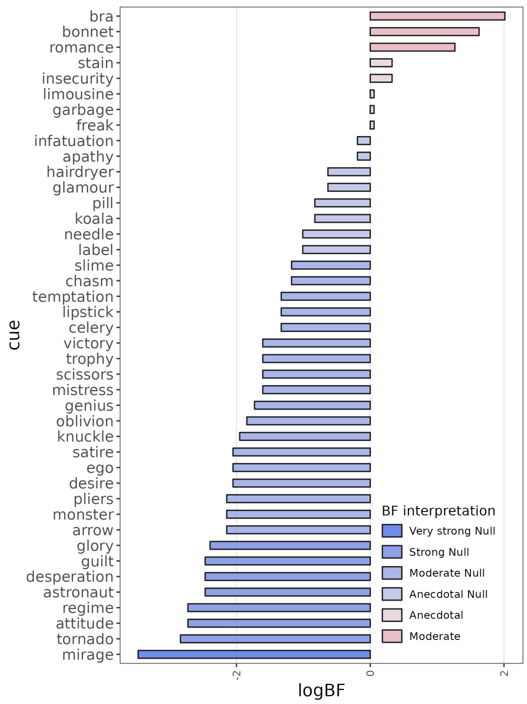

Figure 2: Evidence for gender-specific meaning with BF interpretations according to Jeffreys (1961).

can be attributed to participants' English language proficiency. To do so, we calculated the correlation between an individual's Lextale score and semantic alignment scores toward both English and Chinese speakers. A moderate correlation was found between participants' Lextale scores and the centroid of English speakers, $r(150) = .47, p < .001$. The correlation with the centroid of Chinese speakers was weak, $r(150) = -.27, p = .001$. This suggests that more proficient bilinguals are aligned closer to native English speakers. [2]

Sources of Conceptual diversity

While the semantic alignment analysis shows widespread differences in word meaning, and differences were larger among less proficient speakers, previous work suggests that even within a single language group, meanings can vary to the extent that they could suggest distinct concepts altogether (Martí et al., 2023). To investigate whether words map onto multiple distinct concepts, we used Latent Profile Analysis (LPA) to measure conceptual diversity (see Bauer, 2022, for an introduction). In contrast to the previous semantic distance analysis, where the cue words were represented as a distribution of ratings over associations, clusters (i.e., concepts) are obtained by *grouping participants* who exhibit similar strength ratings for each cue-associate pair.

We employed a Gaussian Mixture Model to quantify distinct concepts by clustering the JAS rating vectors for each individual using the 'mclust' package (Scrucca, Fop, Murphy, & Raftery, 2016). Unlike alternative methods, such as k-means, the Gaussian Mixture Model offers the advantage of automatically determining the number of clusters and detecting evidence within each cluster. To ensure the interpretability of the clustering solution, we randomly sampled an equal number of Chinese and English males and females, resulting in a balanced sample of 4×31 participants across gender and language. While different clusters might map onto systematic variation due to a range of factors, we are primarily interested in establishing to what degree language can explain clustering patterns. However, results for gender will also be included to contextualize language-based variation.

Figure 3 illustrates the clusters in a 2D space using Multidimensional Scaling (MDS). A total of 29 out of 42 words (69.1%) were differentiated into more than 1 cluster, resulting in 77 clusters. The average number of clusters per word was 1.83, with 13 words mapped onto a single cluster, 24 words mapped onto 2 clusters, and 4 words mapped onto 3 clusters and one word onto 4. Similar to the previous section, we used a Bayesian proportion test to determine the evidence for language-specific clusters by comparing the proportion of English and Chinese speakers against chance (which, given the balancing of the groups, was fixed at .50). Of the words that mapped onto multiple clusters, there were 41 (64.1%) clusters with at least anecdotal evidence ($BF_{10} > 1$) for language-based concepts and 32.8% with at east strong evidence. To illustrate the content of these clusters, Figure 4 shows the average ratings for individuals across two clusters of similar size for *satire*. In Cluster 1, which primarily consisted of native English speakers, higher ratings were given to positive and humorous aspects (*comedy*, *humor*, *joke*, *laugh*, *funny*). In contrast, participants in Cluster 2, which primarily included native Chinese speakers, highlighted negative aspects (*mean*, *rude*, *cruel*).

As a baseline comparison, we also investigated how gender (across both languages) was associated with conceptual diversity. Of words that mapped onto multiple clusters, there were 11 (17.2%) words with at least anecdotal evidence ($BF_{10} > 1$) and one word (*koala*, 1.6%) with moderate evidence ($BF_{10} > 3$) for gender-specific concepts. A closer inspection showed that none of these 11 clusters contained words for which we identified gender-meaning differences in the previous section (see Figure 2: *bra*, *bonnet*, *romance*, *stain*, *insecurity*, *limousine*, *garbage*, *freak*).

Finally, we also investigated if lexical factors might explain the conceptual diversity measured through the LPA clustering by correlating the number of clusters with the lexical norms. None of the Spearman correlations between the number of clusters and the words' concreteness, valence and gender stereotypically were significant. Next, we compared the average number of concepts for the discrete factors in the stimuli (concrete vs abstract, emotional vs neutral, and gender-stereotypical vs neutral) to test the hypotheses that abstract, emotional or gender-stereotypical words would map onto a higher number of concepts. We performed a one-sided Wilcoxon rank-sum test for each hypothesis but found

[2] Note that the effect for the Chinese centroid is based on the average scores of all Chinese speakers, which might attenuate the effect size.

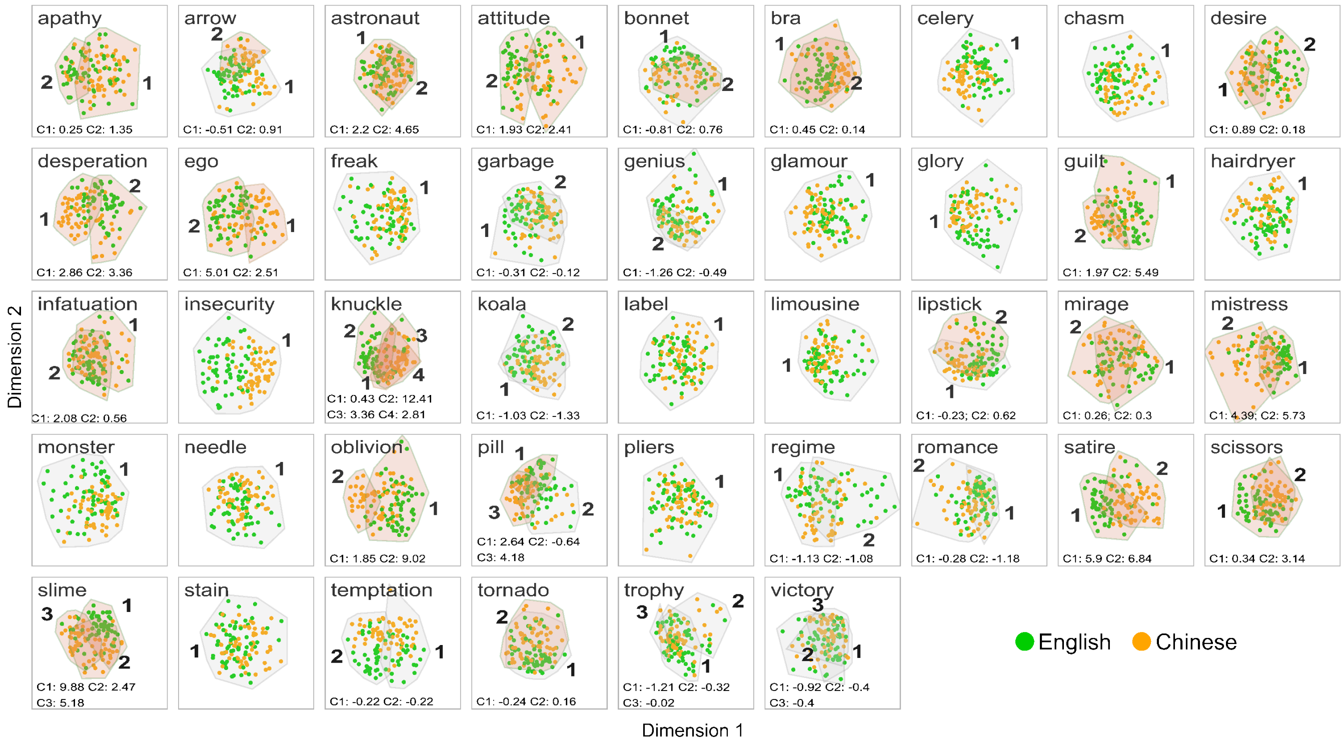

Figure 3: Distribution of English and Chinese Speakers in the Clusters of Associate Ratings for 42 Cue Words. Green dots indicate English native speakers. Orange dots indicate Chinese native speakers. Clusters with evidence of $BF_{10} > 1$ are highlighted, and log-transformed BF for multiple clusters are shown at the bottom of each facet plot.

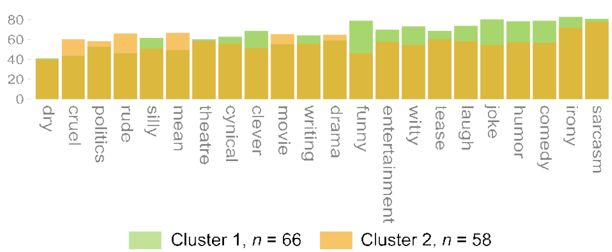

Figure 4: Illustration of conceptual diversity for *satire*. Green bars represent Cluster 1, which predominantly (67.9%) consisted of English speakers. Orange bars represent Cluster 2, which consisted of predominantly (75.0%) Chinese speakers.

no significant differences. As such, we failed to reject the null hypothesis that different types (i.e., lexical factors) of words have similar conceptual diversity.

Discussion

The present study investigated to what degree Chinese non-native speakers of English understood common English words relative to native English speakers. We found evidence for language-specific meanings in all cue words, with strong evidence found for 95% of them. Secondly, we investigated how language contributes to the conceptual diversity of common words. On average, 1.83 concepts per word were identified. Anecdotal evidence for language-specific meanings was found as the source of conceptual diversity in 64% of words with multiple concepts and strong evidence in 32.8% of words. In contrast, gender-specific meaning was not as widespread (8 out of 42 words with at least anecdotal evidence). Similarly, evidence that was at least anecdotal for conceptual diversity related to gender was found for only 17.2% of cases. Lastly, our results did not provide significant evidence that meaning variation depends on how emotional, abstract, or gender-laden a word is, with several concrete, gender and affect-neutral words showing differences across languages and mapping onto multiple concepts.

The current findings provide robust evidence of semantic differences in English word meaning among native English and Chinese speakers, even for common words judged to be familiar among our non-native participants. Theoretically, our study's results on semantic diversity lend support to the Modified Hierarchical Model proposed by Pavlenko (2007), which accommodates language-specific conceptual representations. This is in contrast to the Revised Hierarchical Model by Kroll et al. (2010), which assumes shared conceptual representations across languages.We also propose the Sense Model Finkbeiner, Forster, Nicol, and Nakamura (2004) as a potential explanations of our findings. It could be postulated that, compared to native English speakers, non-native English speakers tend to assign lower ratings to subordinate senses of cue words.

Moreover, our study on conceptual diversity replicated the results for monolingual English speakers (Martí et al., 2023), supporting the existence of multiple concepts associated with single cue words. However, a relatively lower total count of concepts than Martí et al. (2023) was found, who reported 6 to 11 distinct and 6 to 16 concepts for the domains of animals and politicians. Apart from different semantic domains, one possible explanation could be the discrepancy in sample size, as Martí et al. (2023) employed a larger group of 1,000 participants in their JAS task, while our study involved only 152 participants. Hence, Martí et al. (2023) study may have been more likely to detect evidence of conceptual diversity due to the larger sample size. A second potential factor contributing to this discrepancy is the lack of rating standardization, which might also contribute to inflated cluster counts in (Martí et al., 2023). A third factor is related to the different participant pools, where Martí et al. (2023)'s participants were likely to be more heterogeneous as they were recruited through Amazon Mechanical Turk, which might have resulted in increased conceptual diversity.

Our findings also provide new insights into the prevalence of gender-related differences in meaning and concepts. Our study conceptually replicated the findings of De Deyne et al. (2023) and extended the constrained adjective association task to a free word association task. This modification allowed us to include a broader range of parts of speech for the associated words rather than being restricted to adjectives alone. Our findings revealed gender-specific meanings in 8 (19%) of the 42 cue words, which is lower than the results reported by De Deyne et al. (2023), who identified gender-specific meanings for 12 (29%). Conceptual diversity related to gender was also lower, with only 19% of words having at least one gender-related concept in the current study compared to 31% in De Deyne et al. (2023). The lower results could be due to several factors, such as the use of free association tasks or the presence of non-native speakers within both gender groups. Furthermore, when comparing the gender-specific cue words identified in our research with those found in De Deyne et al. (2023)'s study, we observed that only three cue words, *bonnet*, *insecurity* and *bra*, exhibited overlapping gender-specific meanings in both studies. All other gender-specific cue words were unique to one study and not the other. Since our comparison across genders relies on speakers of different languages, our results might have masked gender effects that are more pronounced in homogeneous samples. A follow-up with a larger sample of male and female participants within each language group could shed light on this.

When investigating the role of concreteness in a cross-lingual context, semantic differences between languages were not significantly associated with the concreteness of the cue words. This suggests that abstract words are not necessarily conceptually more diverse than concrete words, contrasting with previous work in Chinese monolinguals where larger meaning differences for abstract concepts were found (Wang & Bi, 2021). Furthermore, our findings did not reveal a significant differentiation between affective words (positive or negative) and neutral words, or gender-laden and neutral words.

Taking a broader perspective, the widespread differences both in terms of (continuous) differences in meaning and the possibility that some of these differences might reflect different concepts highlight the challenge for non-native speakers to understand even common L2 words. This carries practical significance since neglecting these language and cultural distinctions can lead to translation disparities and misunderstandings, especially for classroom-based language learners who may not be well-acquainted with the cultural differences of a second language.

Limitations and Future Directions

While our analyses provide initial insight into what words have different meanings, the stimulus sample size is relatively small, which prevents us from drawing firm conclusions about the role of lexical factors such as concreteness or valence. However, work in progress with a much larger sample of words and participants aims to address this issue, with preliminary results confirming that meaning differences occur for all kinds of words, including concrete and neutral ones. The role of language proficiency and conceptual diversity is a second factor that deserves more attention. The current analysis includes participants at an Australian university, which means that the range of proficiency is somewhat restricted. It is, therefore, likely that conceptual diversity is underestimated when only highly proficient students are included. More generally, learning the meaning of a word is complex because multiple properties or semantic features are likely to be acquired simultaneously. This suggests that the relations between conceptual diversity and proficiency might be non-linear. Finally, the current work has not explicitly considered the role of translation or L1-to-L2 semantic transfer among Chinese speakers. Here also, the functional relationship between semantic alignment or conceptual diversity, on the one hand, and cross-lingual distance might be complex depending on whether a Chinese (near-) translation equivalent is available. Finally, including multiple languages of varying cross-lingual distance might be particularly useful to triangulate whether conceptual diversity reflects L1 meaning transfer or different acquisition rates of specific English connotations and senses and could contribute to the debate about the degree to which meaning is universal across languages or culture/language specific (Thompson et al., 2020).

Conclusion

Native English and highly proficient non-native English speakers understand commonly used English words differently, and this difference was found for all words included in this study. Moreover, our investigation into conceptual diversity demonstrates that most words map onto multiple concepts. While multiple factors contribute to this conceptual diversity, the individuals' language background was found to be one of the major ones. While gender-specific word meanings and concepts were found, the effect of language considerably surpasses that of gender.

References

Bauer, J. (2022). A primer to latent profile and latent class analysis. In *Methods for researching professional learning and development: Challenges, applications and empirical illustrations* (pp. 243–268). Springer.

Capitani, E., Laiacona, M., & Barbarotto, R. (1999). Gender affects word retrieval of certain categories in semantic fluency tasks. *Cortex*, *35*(2), 273–278.

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.

De Deyne, S., Warner, S., & Perfors, A. (2023). Common words, uncommon meanings: Evidence for widespread gender differences in word meaning. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Fairfield, B., Ambrosini, E., Mammarella, N., & Montefinese, M. (2017). Affective norms for italian words in older adults: age differences in ratings of valence, arousal and dominance. *PloS one*, *12*(1), e0169472.

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, *51*, 1–22.

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., … Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517–1522.

Jeffreys, H. (1961). *The theory of probability*. Oxford University Press.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(3), 373–381.

Lemhöfer, K., & Broersma, M. (2012). Introducing lextale: A quick and valid lexical test for advanced learners of english. *Behavior research methods*, *44*, 325–343.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (leap-q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 940–967.

Martí, L., Wu, S., Piantadosi, S. T., & Kidd, C. (2023). Latent diversity in human concepts. *Open Mind*.

Pavlenko, A. (2007). *Emotions and multilingualism*. Cambridge University Press.

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, *89*(4), 344–350.

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, *51*(3), 1258–1270.

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal*, *8*(1), 289.

Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, *4*(10), 1029–1038.

Vivas, L., Montefinese, M., Bolognesi, M., & Vivas, J. (2020). Core features: measures and characterization for different languages. *Cognitive Processing*, *21*(4), 651–667.

Wang, X., & Bi, Y. (2021). Idiosyncratic tower of babel: Individual differences in word-meaning representation increase as word abstractness increases. *Psychological Science*, *32*(10), 1617–1635.