The Contextual Variability of English Nouns: The Impact of Categorical Specificity beyond Conceptual Concreteness

Giulia Rambelli, Marianna M. Bolognesi

University of Bologna giulia.rambelli4@unibo.it, m.bolognesi@unibo.it,

Abstract

Research on conceptual abstraction has investigated the differences in contextual distributions, or "contextual variability," of abstract and concrete concept words (e.g., *love* vs. *cat*). Empirical studies on this topic show that abstract words tend to occur in diverse linguistic contexts, while concrete words are typically constrained within more homogeneous contexts. Nonetheless, these investigations have somewhat overlooked a factor that influences both abstract and concrete concepts: *Categorial Specificity*, which denotes the inclusiveness of a category (e.g., *ragdoll* vs. *mammal*). We argue that more specific words are tied to narrower domains, independently or whether they are concrete or abstract, thus resulting in a diminished degree of contextual variability when compared to generic terms. In this study, we used distributional models to investigate the interplay between contextual variability, concreteness, specificity, and their interaction. Analyzing 676 English nouns, we found that contextual variability is explained by both concreteness and specificity: more specific words have closer contexts, while generic words, whether abstract or concrete, exhibit less related contexts.

Keywords: Cognitive Methods, Semantics, Lexicon

1. Introduction and Related Works

Gaining a comprehensive understanding of the cognitive mechanisms underlying the processing of concrete and abstract meanings remains a pivotal yet unresolved question in cognitive science (Barsalou and Wiemer-Hastings, 2005). One of the most prominent theories explaining the distinctions between these two types is the Context Availability Hypothesis (Schwanenflugel et al., 1988), according to which there are differences in the availability and strength of contextual associations between concrete and abstract words. Specifically, concrete words are deemed "semantically richer" than abstract ones, thereby accounting for their processing advantage, often referred to as the "concreteness effect" (Jessen et al., 2000).

The exploration of the distributional attributes of concrete and abstract concepts and words has followed different routes, employing diverse metrics to investigate how words manifest within contexts. We employ the overarching term "contextual variability" (CV; Hoffman, 2016) to encompass all the proposed metrics of contextual behaviors. Broadly speaking, contextual variability refers to the number of different contexts in which a word is encountered (Johns and Jones, 2022). Previous research on contextual variability has revealed that words referring to concrete concepts tend to appear in a limited but highly similar set of contexts (Hoffman et al., 2013; Hoffman and Woollams, 2015), while abstract concepts exhibit a higher degree of variability across contexts. Several computational studies have delved into conducting a thorough corpusbased analysis to discern the distinctions between

concrete and abstract words (Recchia and Jones, 2012; Frassinelli et al., 2017; Naumann et al., 2018; Frassinelli and Im Walde, 2019; Schulte im Walde and Frassinelli, 2022). These investigations consistently reveal a common pattern: concrete words display a preference for co-occurring with other concrete words, while abstract words tend to co-occur more frequently with other abstract words.

However, previous investigations have concentrated on the divergence between concrete and abstract concepts, disregarding their difference in Categorial Specificity, namely, the level of inclusiveness within the referential category (Bolognesi et al., 2020; Bolognesi and Caselli, 2022). This methodology can be problematic, as it might lead to comparisons between very specific concrete concepts, such as strawberry, and highly generic abstract concepts like knowledge, or very generic concrete concepts like substance and very specific abstract concepts like *bankruptcy*. Crucially, generic and specific words may display distinct contextual distributions: specific words may tend to occur in limited contexts because they refer to precise entities in texts characterized by high-resolution semantics. Conversely, generic words may be found in a broader array of diverse contexts due to their less precise nature, making them adaptable to different contexts. Moreover, generic words may be encountered in texts characterized by low-resolution semantics, thus appearing in a broader range of loosely related contexts.

Contributions This study has a dual purpose: i.) to enhance our understanding of the distinctions between concrete and abstract words and concepts while introducing the previously overlooked factor of

specificity, and ii.) to explore potential variations in contextual variability across languages, particularly between English and Italian, analyzed by Rambelli and Bolognesi (2023). By providing a contextual analysis of words that vary in concreteness while considering specificity, this research contributes significantly to the ongoing discussion regarding how meaning is represented in the human mind. Additionally, it holds promise for improving NLP applications (such as text classification, summarization, simplification, and information retrieval), ultimately enhancing the accuracy and quality of tasks related to language understanding and generation.

2. Materials and Methods

2.1. Data

For this study, we employed words annotated with specificity scores collected using the same method described in Bolognesi and Caselli for Italian by Ravelli et al. (2024). Specificity ratings were obtained for 1,034 words from the ANEW dataset. These words are also annotated for concreteness (ratings gathered by Brysbaert et al., 2014).

To compute contextual variability measures, we built a Distributional Semantic Space (DSM) for English using the word2vecf model (Levy and Goldberg, 2014). Co-occurrences were extracted from a large corpus of English comprising ukWaC (Baroni et al., 2009) and Wikipedia 2018, totaling about 2 billion tokens. The corpora were processed using the CoreNLP pipeline (Manning et al., 2014) for tokenization, lemmatization PoS tagging, and dependency parsing. The DSM was trained on nouns, verbs, and adjectives with frequencies of at least 100. We extracted <target, context> pairs within a window of ± 10 words with a frequency > 20¹. We used the skip-gram algorithm with default settings: no hierarchical softmax, 15 negative samples, and 300 vector dimensions. For experiments, we selected 676 nouns from the dataset attested in the DSM².

2.2. Contextual Variability's Measures

Previous empirical models have yielded a noteworthy insight into the relationship between word abstractness and contextual diversity. Essentially, these models indicate that abstract words tend to appear in a wider array of contexts, while concrete words are typically found in fewer contexts (Recchia and Jones, 2012; Hill et al., 2014). To operationalize the concept of contextual variability – which pertains to how closely a word relates to its contexts – researchers have employed Distributional Semantic Models (DSMs). Following Rambelli and Bolognesi (2023), we computed several measures of contextual variability, organized into two groups.

Neighborhood density indicates how closely a word relates to its paradigmatic neighbors within the distributional space. We used two measures introduced by Schulte im Walde and Frassinelli (2022): Target-Neighbors (TN) similarity, which calculates the average vector-space distance between a target word and its k-nearest neighbors, and Neighbors-Neighbors similarity (NN), which guantifies the average vector-space distance between the k-nearest neighbors of the target word. Conversely, Context Richness explores the syntagmatic contexts in which a word appears, focusing on the strength of the relationship between a target noun and its most associated contexts. We employed measures inspired by Schulte im Walde and Frassinelli (2022), namely, Target-Contexts similarity (TC) and Contexts-Contexts similarity (CC), and the Distributional of Context Richness (DCR) measure proposed by Lenci et al. (2018), which ranks contexts and calculates the mean of scores of the k-top contexts. Finally, we computed **Contextual** Entropy (H; Shannon 1948), which quantifies the informativeness of linguistic contexts surrounding a word. Higher entropy values suggest greater uncertainty in word occurrence given its linguistic contexts (McDonald and Shillcock, 2001 use entropy to model their "Contextual Distinctiveness" measure). Neighborhood density and context richness represent similarities with paradigmatic or syntagmatic contexts, so we maintain their separation to avoid misinterpretations in our analyses.

3. Experiments

Given our 676 selected nouns, we calculated all the CV metrics mentioned above, varying the value of k (5, 10, 20, 50) to understand the impact of different numbers of contexts/neighbors on TC, CC, TN, and NN scores. We ranked contexts using two association measures: Positive Pointwise Mutual Information (Church and Hanks, 1990), used for DCR, and Local Mutual Information (Evert, 2009)³, applied to select contexts for TC and CC.

In the main experiment, we conducted a series of regression analyses⁴ with the objective of uncovering the relationships between CV metrics and

¹Given that a larger window size produced higher R^2 values for several CV metrics (compared to a window size = 2; cf. Rambelli and Bolognesi, 2023), we chose to replicate the study using only a window size of 10.

²Data and scripts available at https://osf.io/mhsdv/?view_only= 2636353fb4d34e80a942540a42f28469.

³LMI is the co-occurrence frequency multiplied by PMI and mitigates bias towards low-frequency events.

⁴Computed in R (v. 3.6.3) with stats package.

Concreteness/Specificity scores. In a more granular breakdown, we performed linear regressions where the contextual variability metric served as the dependent variable, and we considered three independent variables: i) solely the Concreteness score, ii) solely the Specificity score, and iii) the interaction between Concreteness and Specificity. In each model, all predictors were centered. Figure 1 reports the coefficient of determination, denoted as Adjusted R^2 , which signifies the proportion of the overall variability in the dependent variable that the regression model accounts for. Higher R^2 values (indicated by darker colors) suggest that the regression model effectively explains a substantial portion of the variation in context variability.

3.1. Regression Analyses

Concreteness effect Firstly, we observed how each measure is affected by Concreteness. In Rambelli and Bolognesi (2023), concreteness values of Italian words account for only a modest proportion of the variance in contextual variability scores, ranging from 1.3% to 5%. For English, some CV measures are better explained by Concreteness, namely NN and TC ones, which explain over 10% of the variance. Moreover, these two scores strongly correlate (Spearman's ρ =0.38⁵). We could infer that these two measures are complementary: concrete words occur in very similar contexts and with each other, providing similar representations for all words that co-occur in the same scenario. These results confirm Schulte im Walde and Frassinelli (2022): the distributionally most similar context words in relation to a target (TC) depends on the target's concreteness, i.e., the higher this average vector-space similarity is, the more concrete the target words are. However, Concreteness also partially explains NN, while it is the worst neighborhood density for detecting concrete nouns for both previous English experiments. It is still to investigate whether it depends on the dataset design or the DSM over which similarities are computed.

Specificity effect Considering the model with Specificity as an independent variable, we found that NN and TC have a R^2 of around 9-13%, although they are somewhat lower compared to the previous model. Nevertheless, contextual entropy plays a prominent role in accounting for variance. Notably, we observe a negative correlation between Specificity and entropy: as Specificity increases, entropy decreases. This implies that specific words, whether concrete or abstract, are highly expected within their contexts, whereas generic words are more surprising because they appear in a wider range of contexts, leading to higher entropy. This finding aligns with what was reported



Figure 1: Summary of the three groups of regressions (columns) given CV measures as the dependent variable (rows). Cells report Adjusted R^2 values and *p*-values (*:<.05, **:<.01, and ***:<.001).

by Schulte im Walde and Frassinelli (2022), where entropy emerges as a significant predictor for distinguishing the more specific word in a pair.

To illustrate this point, consider the case of pasta and food: they are both concrete words (4.86 and 4.8, respectively) but differ in specificity (ratings: 4.23 and 1.52, respectively), and consequently they have a different association to their contexts (TC 10: .65 vs .49). For instance, pasta has other types of food as top-contexts, such as dish (0.66^6) , sauce (.81), bread (.68), rice (.59), food (.49), salad (.78). Conversely, food is associated heterogeneous words describing events (verbs like eat (.64), find (.29)), nouns loosely related to food (drink (.61), chain (.35), animal (.51), and typical collocations (fast (.22) for fast food). This trend is present, thus less evident, even for abstract words: a specific word like bereavement (c:2.33; s:3.38) is associated with nouns related to grieving (allowance (.41) benefit (.37), support (0.29), loss (.28); grief (0.55)). On the contrary, a generic term like wish (c:.77; s:2.1) highly occurs with less semantically related words (e.g., best (.19), express (.32), have/V (.19), list (.23), grant). This observation supports the findings for Italian nouns and underscores the idea that more generic words tend to occur in a variety of contexts that are not closely tied to the target word, while more specific words exhibit a stronger association with similar contexts.

Interaction effects While CV measures have been investigated to predict Concreteness or Specificity, the interaction between the two metrics was never taken into consideration for English nouns to the best of our knowledge. The results again reveal that TC measures are good metrics, encompassing a

⁵All reported correlations have *p*-values<.001.

⁶Number in parentheses refer to the cosine similarity between the target noun and the context word.



Figure 2: Interaction plot: relationship between Conc and TC_10 for different levels of Spec.



Figure 3: Interaction plot: Concreteness and Entropy for different levels of Spec.

greater portion of variance compared to the two preceding models, as the interplay of these two factors influences them. Furthermore, the R^2 score for the model with entropy exceeds 26%, similar to what was observed in the case of Italian when considering contextual entropy. Figure 2 visually illustrates the interaction effects between Concreteness and Specificity on the TC_10 measure, which quantifies the average similarity between a target word and its top 10 contexts. Notably, a "concreteness effect" (Jessen et al., 2000) is evident for specific terms (orange line): lower concreteness scores correlate with lower context richness (TC), while higher concreteness ratings result in greater context similarity. However, generic words (blue line) tend to have the same score independently of their concreteness. Figure 3 reports the interaction on entropy. Similarly to what was reported for Italian: Generic words, whether concrete or abstract, exhibit high entropy (blue line), indicating they are less expected within the given context. In contrast, specific words (orange line) have a low entropy, with abstract-specific words having lower entropy than concrete-specific words, implying that highly specific and abstract words are more predictable within their context than highly specific and concrete ones.

3.2. Correlations

We computed the correlations among specificity ratings, concreteness norms, and CV measures.

First of all, we observe a high correlation between Concreteness and Specificity (Spearman's ρ =0.71), which is higher than what was observed for Italian, but in line with the correlation coefficient reported by Schulte im Walde and Frassinelli (2022) for a subset of 226 English nouns (Spearman's ρ =.704). Moreover, Concreteness positively correlates with TC 10 and NN 5 (Spearman's ρ = .44 and .42, respectively). Overall, TC and TN measures are highly correlated, indicating that neighborhood density and contextual richness are closely related. Finally, all distributional metrics are positively correlated with Concreteness and Specificity, with an exception for entropy, which is negatively correlated with Specificity (Spearman's ρ = -0.42); it also exhibits no correlation with Concreteness $(\rho = -0.05).$

As an additional investigation, we computed the correlation between Italian and English scores⁷. First, the variables of primary interest, namely Concreteness and Specificity, remain relatively stable between the two languages, with high correlation scores (.824 and .797, respectively). We also computed the correlation between CV measures in Italian and English (Table 1). Among others, entropy shows a high correlation (.746), while the other CV measures have a slight positive correlation, which usually decreases with the increasing number of neighbors/contexts. This observation tells us that entropy is a robust measure of CV independently of language and corpus size. At the same time, further investigations have to be carried out to understand how the other metrics are reliable for this kind of investigation.

3.3. Analysis of Contexts

Finally, we looked at the lexical-semantic properties of the context words, as we believe this could provide useful insight into the differences between words varying in Concreteness and Specificity. We offer below some preliminary considerations for nouns, but the distributions of the contexts would benefit a larger analysis across partof-speech (POS) to identify predominant patterns. Firstly, nouns are the most common POS associated with other nouns (around 50% or more), with verbs and adjectives being less frequent. When ordering words by Concreteness or Specificity, it's evident that more concrete words typically have other nouns as exclusive contexts, although there's no such pattern for specific terms. Furthermore, we examined how concrete the top 10 contexts of target nouns are for various concreteness and specificity values. Considering only the words within Brys-

 $^{^7 \}rm Correlations$ were computed over 586 nouns attested in both Italian and English analyses using Spearman's ρ coefficient.

metric	Spearman's ρ
specificity	0.797
concreteness	0.824
TN_5	0.323
TN_10	0.308
TN_20	0.272
TN_50	0.192
NN_5	0.158
NN_10	0.137
NN_20	0.065
NN_50	0.013
TC_5	0.238
TC_20	0.207
TC_10	0.269
TC_50	0.116
CC_5	0.196
CC_10	0.218
CC_20	0.219
CC_50	0.183
entropy	0.746
DCR	0.442

Table 1: Correlations between Italian and English metrics.

baert's dataset, we found that the concreteness of contexts is higher for more concrete nouns (Spearman's ρ = 0.71; Frassinelli and Im Walde 2019), and also, more specific words tend to have more concrete contexts (Spearman's ρ = 0.62). Similarly, we computed the Specificity values of context words by applying the metric proposed in Bolognesi et al. (2020) (cf. Specificity 3 measure). Correlations are moderately high (Spearman's ρ = .42 for Concreteness, ρ =.46 for Specificity), proving that context words also vary in specificity when a word is more or less concrete or generic. Overall, this is further confirmation that adding the Specificity axis provides a better understanding of the distributional signature of words instead of relying on only Concreteness.

4. Conclusion

These analyses hereby presented provide a more comprehensive view of the relationship between abstraction and contextual variability compared to previous research. Notably, by considering an overlooked aspect of abstraction, *Categorical Specificity*, we have found that contextual variability differences depend on both Specificity and Concreteness. In particular, specific words have well-defined, similar contexts, while generic words, whether abstract or concrete, have broader and more diverse contexts. Unlike in Italian, Concreteness plays a more significant role in explaining noun contextual variability in English. However, Specificity, or the interaction between these two factors, accounts for a more significant portion of the variation in the regression analyses. Additionally, a correlation analysis demonstrated that the measure of entropy is cross-linguistically reliable, while measures computed using similar neighbors or syntagmatic contexts are correlated but more languagedependant. Future research should delve into a more detailed examination of other parts of speech, such as adjectives and verbs. Efforts should also focus on re-running the analysis on Italian and English corpora of comparable size.

Acknowledgment

Funded by the European Research Council, Abstraction project (attributed to Marianna Bolognesi. Grant agreement: ERC-2021-STG-101039777). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

5. Bibliographical References

- Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43:209–226.
- Lawrence W Barsalou and Katja Wiemer-Hastings. 2005. Situating abstract concepts. *Grounding cognition: The role of perception and action in memory, language, and thought*, pages 129–163.
- Marianna Bolognesi, Christian Burgers, and Tommaso Caselli. 2020. On abstraction: decoupling conceptual concreteness and categorical specificity. *Cognitive Processing*, 21(3):365–381.
- Marianna Marcella Bolognesi and Tommaso Caselli. 2022. Specificity ratings for italian data. *Behavior Research Methods*, pages 1–18.
- Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, 46:904–911.
- Kenneth Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29.

- Stefan Evert. 2009. *Corpora and Collocations*, chapter 58. De Gruyter Mouton, Berlin, New York.
- Diego Frassinelli and Sabine Schulte Im Walde. 2019. Distributional interaction of concreteness and abstractness in verb–noun subcategorisation. In *Proceedings of the 13th International Conference on Computational Semantics-Short Papers*, pages 38–43.
- Diego Frassinelli, Daniela Naumann, Jason Utt, and Sabine Schulte Im Walde. 2017. Contextual characteristics of concrete and abstract words. In *IWCS 2017—12th International Conference on Computational Semantics—Short papers.*
- Felix Hill, Anna Korhonen, and Christian Bentz. 2014. A quantitative empirical analysis of the abstract/concrete distinction. *Cognitive science*, 38(1):162–177.
- Paul Hoffman. 2016. The meaning of 'life' and other abstract words: Insights from neuropsychology. *J. Neuropsychol.*, 10(2):317–343.
- Paul Hoffman, Matthew A Lambon Ralph, and Timothy T Rogers. 2013. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior research methods*, 45:718–730.
- Paul Hoffman and Anna M Woollams. 2015. Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance*, 41(2):385.
- Frank Jessen, Reinhard Heun, Michael Erb, D-O Granath, Uwe Klose, Andreas Papassotiropoulos, and Wolfgang Grodd. 2000. The concreteness effect: Evidence for dual coding and context availability. *Brain and language*, 74(1):103–112.
- Brendan T Johns and Michael N Jones. 2022. Content matters: Measures of contextual diversity must consider semantic content. *Journal of Memory and Language*, 123:104313.
- Alessandro Lenci, Gianluca E Lebani, and Lucia C Passaro. 2018. The emotions of abstract words: A distributional semantic analysis. *Topics in cognitive science*, 10(3):550–572.
- Omer Levy and Yoav Goldberg. 2014. Dependencybased word embeddings. In *Proceedings of the* 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 302–308.
- Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and

David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.

- Scott A. McDonald and Richard C. Shillcock. 2001. Rethinking the Word Frequency Effect: The Neglected Role of Distributional Information in Lexical Processing. *Language and Speech*, 44(3):295–322.
- Daniela Naumann, Diego Frassinelli, and Sabine Schulte im Walde. 2018. Quantitative semantic variation in the contexts of concrete and abstract words. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 76–85, New Orleans, Louisiana. Association for Computational Linguistics.
- Giulia Rambelli and Marianna M Bolognesi. 2023. Contextual variability depends on categorical specificity rather than conceptual concreteness: A distributional investigation on italian data. In Proceedings of the 15th International Conference on Computational Semantics (IWCS 2023).
- Andrea Ravelli, Marianna Bolognesi, and Tommaso Caselli. 2024. Specificity Ratings for English Data. Manuscript submitted for publication.
- Gabriel Recchia and Michael N Jones. 2012. The semantic richness of abstract concepts. *Frontiers in human neuroscience*, 6:315.
- Sabine Schulte im Walde and Diego Frassinelli. 2022. Distributional measures of semantic abstraction. *Frontiers in artificial intelligence*, 4:206.
- 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.
- Claude E Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

A. CV metrics

We report below the formula implemented for all CV measures (except for Entropy which was introduced in the main text).

• **TN**: the average vector-space distance between *t* and its *k* nearest neighbors.

$$TN(t) = \frac{1}{k} \sum_{i=1}^{k} similarity(t, i)$$
 (1)

• **NN**: the average vector-space distance between the *k* nearest neighbors of *t*.

$$NN(t) = \frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{k} similarity(i,j) \quad (2)$$

where $i \neq j$

• **TC**: the average vector-space distance between *t* and its *k* top contexts.

$$TC(t) = \frac{1}{k} \sum_{c=1}^{k} PPMI(t, c_i)$$
 (3)

• **CC**: the average vector-space distance between the *k* top contexts of *t*.

$$CC(t) = \frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{k} similarity(i, j)$$
 (4)

where $i \neq j$

• **DCR**: the mean of the PPMI scores of the *k* top contexts of the target noun *t*.

$$DCR(t) = \frac{1}{k} \sum_{i=1}^{k} PPMI(t, i)$$
 (5)

B. Descriptive Analysis

We report below the mean and standard deviation of all CV metrics computed.

metric	mean	stdev
TN_5	0.615	0.064
TN_10	0.589	0.060
TN_20	0.562	0.057
TN_50	0.525	0.055
NN_5	0.565	0.092
NN_10	0.542	0.081
NN_20	0.523	0.078
NN_50	0.501	0.076
TC_5	0.435	0.103
TC_10	0.415	0.084
TC_20	0.390	0.070
TC_50	0.350	0.058
CC_5	0.325	0.083
CC_10	0.307	0.059
CC_20	0.288	0.045
CC_50	0.262	0.031
DCR	165.796	1353.744
entropy	8.785	1.909

Table 2: Descriptive statistics of CV measures.