Contextual Variability Depends on Categorical Specificity rather than Conceptual Concreteness: A Distributional Investigation on Italian data

Giulia Rambelli
University of Bologna
giulia.rambelli4@unibo.it

Marianna M. Bolognesi
University of Bologna
m.bolognesi@unibo.it

Abstract

A large amount of literature on conceptual abstraction has investigated the differences in contextual distribution (namely *contextual variability*) between abstract and concrete concept words (*joy* vs. *apple*), showing that abstract words tend to be used in a wide variety of linguistic contexts. In contrast, concrete words usually occur in a few very similar contexts. However, these studies do not take into account another process that affects both abstract and concrete concepts alike: *specificity*, that is, how inclusive a category is (*ragdoll* vs. *mammal*). We argue that the more a word is specific, the more its usage is tied to specific domains, and therefore its contextual variability is more limited compared to generic words.

In this work, we used distributional semantic models to model the interplay between contextual variability measures and i) concreteness, ii) specificity, and iii) the interaction between the two variables. Distributional analyses on 662 Italian nouns showed that contextual variability is mainly explainable in terms of specificity or by the interaction between concreteness and specificity[1]. In particular, the more specific a word is, the more its contexts will be close to it. In contrast, generic words have less related contexts, regardless of whether they are concrete or abstract.

1 Introduction

In the study of lexical semantic representation, an extensive debate focuses on explaining the differences between words referring to concrete and abstract concepts. According to the *Dual Coding Theory* (Paivio, 1991), concrete words are represented in two different systems, one language-based and one image-based, while abstract words are primarily or exclusively represented in the former system. The *Context Availability Hypothesis* (Schwanenflugel, 2013) instead argues that all word meanings are represented in a single verbal code, but concrete words have stronger and denser associations to contextual knowledge than abstract ones. Both theories agree on two points: i) the meaning of abstract words is essentially acquired via language, for instance, through distributional statistics extracted from the linguistic input, and ii) concrete words are "semantically richer" than abstract ones, thereby explaining their processing advantage, the so-called *concreteness effect* (Jessen et al., 2000).

The investigation of the distributional properties of concrete and abstract concepts and words has taken different paths, implementing different metrics to measure how words behave in context (see Section 2.1). We hereby use the general term *contextual variability* (Hoffman, 2016) as an 'umbrella' that includes all proposed metrics of contextual behaviors, described in the next section. Overall, the previous works on contextual variability showed that words referring to concrete concepts occur in a few but very similar syntagmatic contexts, depending on the fact that their meanings are tied to a fixed class of objects or events in the environment. On the other hand, abstract concepts are characterized by a greater degree of variability across contexts, commonly attributed to their association with less well-defined, intangible experiences or properties.

Notwithstanding, it is worth noting that prior investigations have mainly focused on the divergence between concrete and abstract concepts, while potentially overlooking any discrepancies in specificity, that is, the level of inclusivity in the referential category. This can be problematic because it may lead to comparisons between very specific concrete concepts like *muffler* and very generic abstract concepts like *manner*, or very generic concrete concepts like *substance* and very specific abstract concepts like *sorrow*. Crucially, generic and specific

[1] Data available at https://osf.io/2qm5e/?view_only=fce6b4bb895a41658ed97512afa65ae3

words may have different contextual distributions: specific words may tend to be used in limited sets of contexts because they denote precise entities occurring in texts characterized by high-resolution semantics. Conversely, generic words may be used in a wider range of diverse contexts because they are less precise and, therefore, more easily applicable to different contexts; generic words may occur in texts characterized by low-resolution semantics and, therefore, may occur with a wider range of shallowly-related contexts.

With the present study, we tackle the following questions:

- How does concreteness explain the variation in contextual distributions of nouns?
- How does specificity explain the variation in contextual distributions of nouns?
- How does the interaction between concreteness and specificity explain the variation in contextual distributions of nouns?

These questions are addressed through a series of regression studies in which the concreteness ratings Montefinese et al. (2014) and specificity ratings Bolognesi and Caselli (2022) of 662 Italian nouns are modeled with a set of corpus-based indices representing their context variability.

2 Related works

2.1 Operationalizations of Contextual Variability

When investigating how concrete and abstract concepts are processed in the mind, researchers have endeavored to relate such differences to the differences between the contexts of occurrence (a.k.a. *contextual variability*) of concrete and abstract words (Hoffman, 2016, for a review).

The Context Availability hypothesis, for instance, notes that concrete words tend to have more robust and intricate contextual associations than abstract ones. This notion is supported by Schwanenflugel and Shoben (1983) 's early research, which found that speakers find it easier to imagine a context for concrete words compared to abstract words. Schwanenflugel et al. demonstrated that when an explicit context was provided for concrete and for abstract words alike, the processing advantage of concrete over abstract words disappeared. The authors concluded that abstract words were more difficult to process because participants struggled to place them in a meaningful context, but this difficulty was reduced when an explicit context was provided.

Hoffman et al. (2013) employed the term *semantic diversity* to describe the average similarity between the contexts in which a word appears. They discovered that concrete words are used in a limited, closely interconnected set of contexts. For instance, the term "spinach" typically occurs only in contexts related to cooking and eating which are similar to one another. On the other hand, abstract words (e.g., "life") are used in a more diverse range of unrelated contexts, resulting in high semantic diversity values. Moreover, Recchia and Jones (2012) introduced two contextual measures related to abstract and concrete concepts. The first measure, *contextual dispersion* (CD), refers to the number of different content areas (or domains) in which a word appears, as proposed by Pexman et al. (2008). The second measure is the *number of semantic neighbors* (NSN), which measures the number of words that appear within a particular radius of a high-dimensional semantic space. The authors found that NSN is higher for abstract than for concrete words, and this peculiarity facilitated the processing of abstract concepts in lexical decision tasks.

Overall, cognitive studies tend to indicate that abstract words are more likely to be used in a wider variety of linguistic contexts, shallowly related to the target word. Concrete words tend to be used in tighter networks of similar contexts, and this may facilitate their retrieval.

2.2 Computational Models of Abstraction

In the last decade, several computational models have been suggested to automatically validate the cognitive assumptions about the contextual difference between abstract and concrete concepts.

Similarly to Recchia and Jones (2012), Hill et al. (2014) quantitatively analyzed the different patterns of association for words varying in concreteness, providing possible cognitive underpinnings for the differences observed. The authors showed that abstract concepts occur in a broader range of contexts and are organized according to associative principles; concrete concepts instead have few specific contexts of occurrence, and they tend to be organized according to (semantic) similarity principles. Recently, Frassinelli et al. (2017) investigated the degree of concreteness of co-occurring contexts for concrete and abstract English words. They built a vector space model for nouns from the Brysbaert et al. (2014) concreteness norms; to retain concreteness scores of contexts and distributional neighbors, they restricted the vocabulary to nouns attested in the dataset (that is, they built a symmetric co-occurrence matrix in which all targets and context words are from concreteness norms). The authors reported that the more a noun is concrete, the more it tends to appear with other concrete nouns and has a more extensive range of concreteness scores; on the contrary, the more a word is abstract, the more it occurs with other abstract words. While this outcome aligns with multiple studies in the literature, the methodological choice of restricting the number of contexts to the words attested in Brysbaert et al. (2014) may have biased the actual distributional pattern of these words.

Working on Italian, Lenci et al. (2018) observed that abstract words, which according to some studies tend to be characterized by a heavier emotional load compared to concrete words (Vigliocco et al., 2014, i.a.) tend also to co-occur with contexts with an overall higher emotive load. This has been observed based on affective statistical indices estimated as distributional similarity with a restricted number of seed words strongly associated with a set of basic emotions. This study provides additional empirical evidence to support the tendency for more concrete words to be associated with higher contextual richness. Overall, previous studies indicated that concrete words tend to have less diverse but more compact and strongly associated distributional neighbors than abstract words.

While a variety of computational models have been focusing on the contextual properties of concrete and abstract words, there are virtually no computational models focused on the contextual variability of specific and generic words due to the challenges associated with comparing these two variables. One major obstacle is the lack of human ratings available for measuring specificity. Notably, Schulte im Walde and Frassinelli (2022) offer a unique exception to this trend. The authors tested how various distributional measures represent abstract-concrete and general-specific word pairs (represented as hypernym-hyponym pairs from WordNet, Miller and Fellbaum (1991)). Analyses revealed that the distributional similarity of contextual words surrounding a target (i.e., neighborhood density) predicts word concreteness: the higher the similarity, the more concrete the word tends to be, albeit this effect is more pronounced for nouns than for verbs. Nevertheless, this measure is not useful for correctly predicting the specificity of a word, which depends on frequency and word entropy. To the best of our knowledge, they are the first to include both Concreteness and Specificity in this type of investigation. However, there are two limits to this approach. First, as mentioned above, they operationalized word specificity as a binary property (rather than a continuous variable) extracted from WordNet. Arguably, such binary distinction does not capture the fine-grained information encoded in a continuous variable. In a second stance, the authors keep concreteness and specificity separated without considering the interaction between the two variables in relation to their context variability.

3 Materials and Methods

3.1 Concreteness and Specificity datasets

For our study, we employed the Bolognesi and Caselli (2022) dataset (henceforth, **BC**), a collection of human-generated specificity ratings for 1049 Italian words. Specificity ratings were collected online adopting the Best-Worst Scaling method (Louviere et al., 2015); given 4-word tuples (belonging to the same POS), participants had to select the most specific and the least specific word within each tuple. The words used to collect specificity ratings with this methodology are the same used to collect concreteness ratings by Montefinese et al. (2014). Bolognesi and Caselli investigated the relation between human-generated concreteness and specificity ratings and reported a low positive and significant correlation of 0.316 (Spearman correlation coefficient; $p < 0.05$), corresponding to an R^2 of 0.1. This result is evidence that Concreteness and Specificity capture different aspects of abstraction, which are only partially correlated with one another.

The entire BC dataset contains 771 nouns, 220 adjectives, and 59 verbs. Our study focused only on nouns, the larger group among the three parts of speech (Figure 1).

3.2 Italian Distributional Semantic Spaces

For our experiment, we built a Distributional Semantic Space (DSM) for Italian words. We extracted the textual information from La Repubblica (Baroni et al., 2004) and itWaC (Baroni et al.,

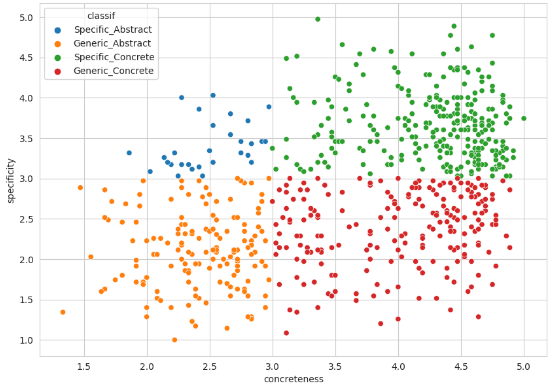


Figure 1: Distribution of the 662 nouns used in the analysis. To approximate the four prototypical types of words different colors are hereby used, although concreteness and specificity have been analyzed as continuous and not as categorical variables.

2009), two pos-tagged and dependency-parsed corpora of Italian. Specifically, we selected a list of nouns, verbs, and adjectives (lemmas used as contexts) with a frequency ≥200 and collected their co-occurrences within a 2- and 10-word symmetric window centered on the target word, which was a noun. We filtered out <target, context> pairs with a frequency of less than 20[2]. The resulting co-occurrence counts were used to i) extract the most associated contexts of a word, using Positive Pointwise Mutual Information (PPMI[3]) score, and ii) built a count-based matrix[4] with PPMI weights and reduced it to 300 dimensions by applying the Singular Value Decomposition (SVD) transformation (Landauer and Dumais, 1997). While we are aware that there are more recent and sophisticated methods, we rely on more stable and explicable representations for the aim of this investigation. We obtained two semantic spaces depending on the context window: **ITAw2** selects nearby words (±2 lemmas surrounding the target word) and contains 19,054 lemmas; **ITAw10** considers a wide contextual window (±10 words) and includes 65,532 lemmas. ITAw10 covers most of the nouns of the BC dataset (754/771), while ITAw2 includes only 662 nouns.

We performed qualitative analyses of the top contexts (CX) and nearest neighbors (NN) for words exemplifying the four prototypical configurations of concreteness and specificity: *abitazione* ('house'; generic concrete), *ambulanza* ('ambulance'; specific concrete), *fantasia* ('fantasy'; generic abstract), and *bancarotta* ('bankrupt'; specific abstract). Tables 1,2, 3, and 4 report the top neighbors (NNs) ordered by cosine similarity, and the top contexts (CXs) ranked by their PPMI with the target word. Comparing the values reported in the tables reveals differences in the contexts extracted using different window sizes. As expected, verbs and adjectives are the most associated contexts within a ±2-word window. Considering a larger context, top contexts are mostly nouns for concrete words (*abitazione*, 'house' and *ambulanza*, 'ambulance'; Table 1 and 2); some verbs are however highly associated to abstract words (*fantasia*, 'fantasy' and *bancarotta*, 'bankrupt'; Table 3 and 4). While the contexts selected are pretty different, the resulting spaces are coherently similar: the neighbors produced by the two spaces overlap a lot, specifically for *abitazione* ('house'; Table 1) and *bancarotta* ('bankrupt'; Table 4). However, similarity scores are considerably lower for ITAw2, indicating that the space is less dense than ITAw10, probably because of the lower number of lemmas and occurrences used to build the DSM.

3.3 Distributional Measures of Contextual Variability

The outcome provided by previous empirical models is that the more abstract a word is, the higher the number of contexts in which it occurs. Conversely, the more concrete a word is, the lower should be the number of its contexts. As introduced above, several computational measures have been proposed to operationalize contextual variability, i.e., how close a word and its contexts are, by relying on DSMs. Given the variety of formulas and terminology, we decided to re-implement previous measures of contextual variability, distinguishing between two subgroups: neighborhood density and contextual richness.

Neighborhood density quantifies how dense the distributional space is near a target word, that is, how close its paradigmatic neighbors are. Looking at a different angle, the higher the average similarity between a word and its neighbors means that many words have a similar contextual distribution. Following Schulte im Walde and Frassinelli (2022), we provide two measures of neighborhood density, Target-Neighbors (TN) similarity and Neighbors-

[2]We tested different values for the filter hyper-parameters and selected the combination that best balances coverage with parser noise.

[3]This is the standard Pointwise Mutual Information, but with negative values raised to 0.

[4]We employed DISSECT toolkit (Dinu et al., 2013).

CX				NN			
w2		w10		w2		w10	
dibire-v 'adhibit-v'	10.82	*censimenti-n* 'census-n'	10.07	*appartamento-n* 'apartment-n'	0.87	*alloggio-n* 'lodging-n'	0.8
perquisire-v 'search-v'	10.44	*furti-n* 'thefts-n'	9.34	*alloggio-n* 'lodging-n'	0.7	*appartamento-n* 'apartment-n'	0.79
irruzione-n 'raid-n'	9.17	*enfiteusi-n* 'emphyteusis-n'	9.16	*edificio-n* 'building-n'	0.63	*fabbricato-n* 'building-n'	0.78
perquisizione-n 'search-n'	8.80	*pertinenziali-n* 'appurtenant-n'	8.97	*immobile-n* real 'estate-n'	0.61	*abitativo-a* 'housing-a'	0.76
lussuoso-a 'luxurious-a'	8.54	*sfitto-a* 'vacant-a'	8.85	*villa-n* 'villa-n'	0.6	*condominio-n* 'condominium-n'	0.74
situare-v 'situate-v'	8.24	*unifamiliare-a* 'single-family-a'	8.68	*albergo-n* 'hotel-n'	0.58	*edificio-n* 'building-n'	0.72

Table 1: Top 6 contexts (CX) and nearest neighbors (NN) of *abitazione* ('house'; spec:2.2, conc:4.63).

CX				NN			
w2		w10		w2		w10	
sirena-n 'siren-s'	12.28	*automedica-n* ambulance 'car-s'	14.95	*pullman-n* 'bus-s'	0.66	*autoambulanza-n* ambulance 'car-s'	0.86
autista-n 'driver-s'	11.20	*barellieri-n* 'stretcher_bearers-n'	14.93	*trafelato-a* 'breathless-a'	0.62	*soccorrere-v* 'rescue-v'	0.81
attrezzare-v 'equip-v'	10.51	*suem-n* 'Medical Service acronym'	13.52	*taxi-n* 'taxi-n'	0.61	*pompiere-n* 'firefighter-n'	0.8
caricare-v 'load-v'	9.86	*bonura-n* -	13.22	*autoambulanza-n* 'ambulance-n'	0.61	*elisoccorso-n* 'helicopter-n'	0.78
trasportare-v 'transport-v'	9.73	*voltolini-n* 'private ambulance service'	12.84	*barella-n* 'stretcher-n'	0.6	*soccorso-n* 'rescue-n'	0.78
croce-n 'cross-n'	8.85	*elisoccorso-n* 'helicopter_ rescue-n'	12.45	*autobus-n* 'bus-n'	0.59	*soccorritore-n* 'rescuer-n'	0.78

Table 2: Top 6 contexts (CX) and nearest neighbors (NN) of *ambulanza* ('ambulance'; spec: 4.14, conc:4.75).

CX				NN			
w2		w10		w2		w10	
inventivo-a 'inventive-a'	12.37	*juvenilia-n* 'juvenilia-n'	11.48	*immaginazione-n* 'imagination-n'	0.7	*immaginazione-n* 'imagination-n'	0.81
fervido-a 'fervid-a'	12.21	*hamill-n* 'hamill-n'	10.68	*invenzione-n* 'invention-n'	0.58	*fantastico-a* 'fantastic-a'	0.78
stuzzicare-v 'tease-v'	12.00	*sbizzarrire-v* 'indulge-v'	10.27	*intelligenza-n* 'intelligence-n'	0.54	*emozione-n* 'emotion-n'	0.76
guizzo-n 'leer-n'	11.12	*solleticare-v* 'tickle-v'	9.57	*immaginario-n* 'imaginary-n'	0.54	*fascino-n* 'charm-n'	0.76
scatenato-a 'unbridled-a'	10.83	*pindarico-a* 'pindaric-a'	9.21	*estro-n* 'whimsical-n'	0.52	*passione-n* 'passion-n'	0.75
sfrenato-a 'unbridled-a'	10.56	*trezzano-n* 'trezzano-n'	9.13	*passione-n* 'passion-n'	0.5	*invenzione-n* 'invention-n'	0.75

Table 3: Top 6 contexts (CX) and nearest neighbors (NN) of *fantasia* ('fantasy'; spec:1.62, conc: 1.66).

CX				NN			
w2		w10		w2		w10	
fraudolento-a 'fraudulent-a'	15.03	*fraudolento-a* 'fraudulent-a'	13.51	*falso-n* 'false-n'	0.8	*concussione-n* 'concussion-n'	0.88
orlo-n 'hemming-n'	11.53	*pluriaggravato-a* 'aggravated-a'	13.29	*peculato-n* 'embezzlement-n'	0.79	*peculato-n* 'embezzlement-n'	0.86
concorrere-v 'concur-v'	10.02	*orlo-n* 'hem-n'	10.78	*appropriazione-n* 'embezzlement-n'	0.78	*fraudolento-a* 'fraudulent-a'	0.85
concorso-n 'conspiracy-n'	9.36	*crac-n* 'crac-n'	9.71	*concussione-n* 'concussion-n'	0.76	*aggiotaggio-n* 'agiotage-n'	0.84
truffa-n 'fraud-n'	7.90	*bancarotta-n* 'bankruptcy-n'	9.64	*ricettazione-n* 'fencing-n'	0.75	*crac-n* 'cracking-n'	0.82
falso-n 'forgery-n'	7.25	*delinquere-v* 'delinquency-v'	9.36	*truffa-n* 'swindling-n'	0.75	*truffa-n* 'fraud-n'	0.81

Table 4: Top 6 contexts (CX) and nearest neighbors (NN) of *bancarotta* ('bankrupt'; spec: 4, conc: 2.27).

Neighbors similarity (NN):

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

Vector-space distance is computed as the cosine similarity between two word vectors.

Conversely, **context richness** looks at the syntagmatic contexts in which a word occurs. It considers the strength of a target noun with its most associated contexts and looks at their respective similarity (similar to semantic diversity). In this case, the highest the value, the more the top contexts have similar vectorial representations, so they refer to similar objects and events; on the contrary, lower scores represent a high variability in the contexts. We implemented several measures of context richness. Target-Contexts similarity (TC) and Contexts-Contexts (CC) similarity are derived from Schulte im Walde and Frassinelli (2022), while *Distributional of Context Richness* (DCR) index was proposed by Lenci et al. (2018):

- **TC**: the average vector-space distance between *t* and its *k* top contexts.
- **CC**: the average vector-space distance between the *k* top contexts of *t*.
- **DCR**: the mean of the PPMI scores of the *k* top contexts of the target noun *t*.

Additionally, we computed the contextual entropy, or average information content (Shannon, 1948), which is a classic measure in computational linguistics and is used as an estimate of context informativeness. The assumption is that the higher the entropy, the more uncertain a word is, or a word is less expected given the linguistic contexts. This measure has been previously introduced as a measure of hypernymy prediction (Santus et al., 2014; Shwartz et al., 2017). We calculated the word entropy (H) considering all the probability between a word and the contexts selected to create the vector space:

$$\mathbf{H}(w) = -\sum_{c} p(c|w) * log_2(p(c|w)) \qquad (1)$$

where $p(c|w)$ is obtained through the ratio between the frequency of $< w, c >$ and the total frequency of w.

	ITAw10		ITAw2	
	M	St.dev	M	St.dev
TN_5	**0.771**	0.069	**0.648**	0.133
TN_10	0.741	0.069	0.610	0.133
TN_20	0.706	0.069	0.566	0.131
TN_50	0.651	0.067	0.498	0.122
NN_5	0.694	0.110	0.609	0.214
NN_10	0.653	0.104	0.558	0.216
NN_20	0.606	0.099	0.496	0.210
NN_50	0.535	0.091	0.392	0.182
TC_5	<u>0.457</u>	0.174	0.239	0.171
TC_10	0.433	0.159	0.225	0.158
TC_20	0.406	0.145	0.208	0.144
TC_50	0.364	0.133	0.191	0.131
CC_5	0.434	0.206	<u>0.277</u>	0.251
CC_10	0.392	0.171	0.231	0.200
CC_20	0.351	0.134	0.191	0.154
CC_50	0.306	0.100	0.154	0.113
DCR	6.009	2.158	5.979	3.476
H	4.783	1.003	4.641	1.065

Table 5: Descriptive statistics of CV measures.

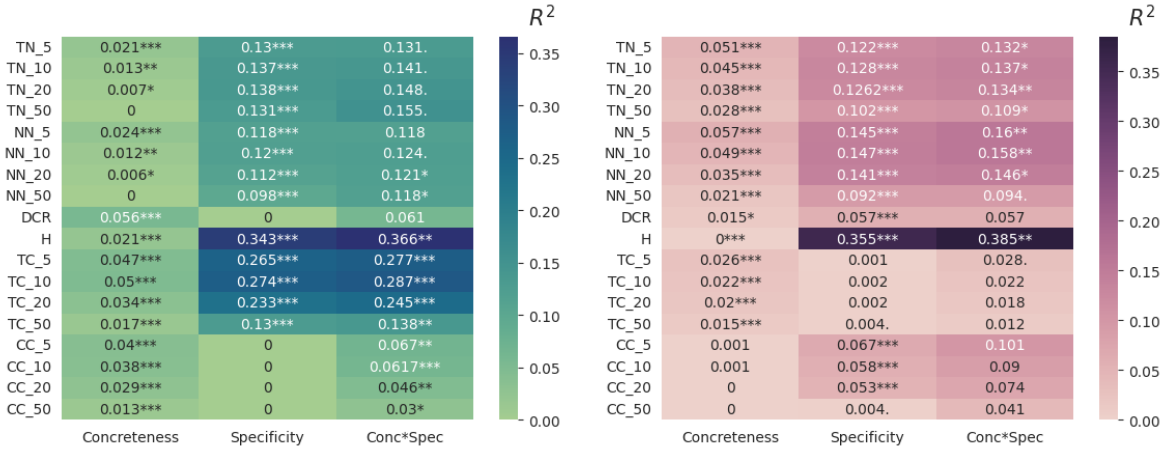
Neighborhood density and context richness are complementary aspects of contextual variability; however, we keep them separated to avoid theoretical and methodological misinterpretations. Formulas are reported in Appendix A.

4 Experimental investigations

Given the 662 nouns attested both in ITAw2 and ITAw10 spaces, we computed all the contextual variability metrics introduced above. We performed the computation with different values of *k* (5, 10, 20, 50) to see how many contexts/neighbors influence the overall score. Table 5 summarizes all computed measures' mean and standard deviation. We observed that the higher the number of contexts/neighbors we select, the lower the overall mean. Moreover, the DCR metric has a high standard deviation, indicating that PPMI scores are not well distributed. The low PPMI scores could be the cause of this issue (see the qualitative analyses above), probably a consequence of the small dimension of the corpora used to extract co-occurrences. This issue is also partially reflected in the entropy measure, with a standard deviation of around 1.

Subsequently, we ran a series of regression analyses[5] aimed at understanding the relations between contextual variability metrics and concreteness/specificity scores. In detail, we ran linear regressions having a context variability metric as the dependent variable; as the independent variable, we consider i) only the Concreteness score, ii) only the Specificity score, and iii) the interaction between

[5]We ran linear models in R (v. 3.6.3) with `stats` package.



(a) ITAw10 (b) ITAw2

Figure 2: Summary of the linear models using Concreteness, Specificity, Concreteness*Specificity as independent variables, and various context density measures as the dependent variable. Cells report Adjusted R^2 values and p-values. '.'=$p < 0.1$, *=$p < .05$, **=$p < .01$, and ***=$p < .001$.

Concreteness and Specificity. The results of the models are reported in Figure 2. The values in the cells correspond to the coefficient of determination R^2, which represents the proportion of the total variation in the dependent variable y accounted for by the regression model. Values of R^2 closer to 1 (darker colors) imply that the regression model explains a large portion of the variance in context variability.

4.1 Main study

The analysis below focuses on interpreting the distributional measures of contextual variability computed on the larger vector space, that is, ITAw10 (Figure 2a).

Concreteness effects Linear models with Concreteness as the independent variable are generally significant, but Concreteness ratings only explain between 1.3% and 5% of contextual variability scores (left column). This outcome reveals that **contextual variability metrics vary as a function of concreteness, but the effect of concreteness on contextual variability is not very high**.

Specificity effects Conversely, Specificity explains the variability of contextual variability values (middle column): TN and NN neighborhood density (around 11-13%), TC context richness (27%), and entropy (34%). However, it does not explain CC metrics. In detail, Specificity explains most of the TC_10 and entropy variance, achieving the highest R^2 scores. The scatterplot in Figure 3 reveals a positive correlation between the two scores (Spearman's ρ = 0.516, $p < 0.001$). Vice versa, entropy is negatively correlated with Specificity (Spearman's ρ = -0.617, $p < 0.001$): the lower the entropy, the higher the Specificity of a word (Figure 4). The two measures reflect the same situation that we can interpret as follows: **more specific words occur in similar contexts**, so they are strongly related to one another, and the word is more expected. Contrariwise, **more generic words are used in a variety of contexts that are not tightly bonded to the target**, so a word is more uncertain for the given context.

We performed a qualitative analysis to corroborate the observed trend. Let us consider the contexts of <u>hamburger</u> (spec: 4.5, conc: 4.1, TC_10: 0.7), a very specific and concrete word. Its contexts are highly similar, and all indicate other kinds of food, such as *ketchup-n*, *patatina-n* ('fries'), *polpetta-n* ('meatball'), *panino-n* ('sandwich'), *manzo-n* ('beef'). Besides, abstract words with high specificity scores have similar associated contexts. Given <u>collera</u> ('rage'; spec: 2.9, conc: 2.8, TC_10:0.71), its contexts are other kinds of emotions, like *lussuria-n* ('lust'), cupidigia-n *(' cupidity')*, *insaziabile-a* ('voracious l'), *brama-n* ('eagerness '), *avidità-n* ('greed').

On the contrary, generic words (i.e., with a low Specificity score) have more heterogeneous contexts, causing a drop in the TC values. For instance, <u>acqua</u> ('water') is concrete but also quite generic (spec: 2.7, conc: 4.7, TC_10: 0.04), and this is reflected in the variety of less related contexts, such as *canaletti-n* ('channels'), *cascatelle-n*

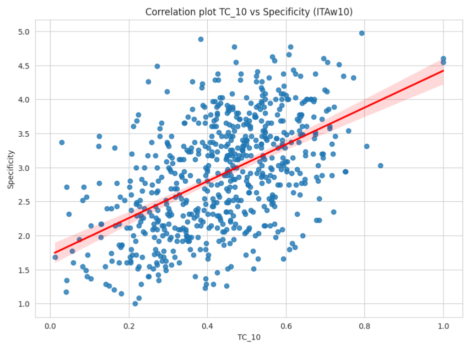


Figure 3: Correlation plots between Specificity and TC_10 measure computed in the ITAw10 space.

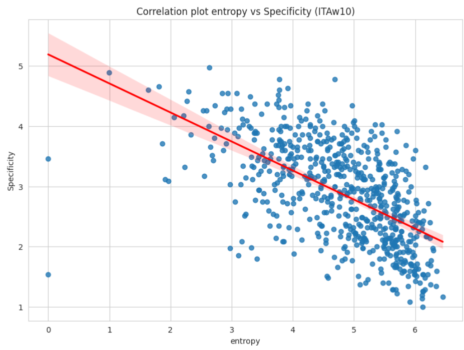


Figure 4: Correlation plots between Specificity and entropy (H) measure computed in the ITAw10 space.

(‘cascade’), *gocciolina-n* (‘drip’), *refrigeratore-n* (‘chiller’), *rigonfiare-v* (‘swell’). Similarly, *tempo* (‘time’; spec: 1.6, conc: 1.6, TC_10: 0.05) has contexts related to the weather, time-traveling, verbal mode, rhythm, and epoch: *viaggiatori-n* (‘traveler’), *zeitgeist-s, trapassato-a* (‘past-tense’), *tiranno-a* (‘tyrant’), *tiranneggiare-v* (‘tyranny’).

It is worth noticing that verbs are more associated with general contexts than specific ones. Qualitative analysis reveals that the difference in the contextual distribution does not overlap with the distinction between abstract and concrete nouns: **Contexts vary depending on the specificity of a word, and this phenomenon is independent of their concreteness**.

Interaction effects Finally, we investigated the interaction between Specificity and Concreteness (right column). Similar to the Specificity models, the interaction explains TC_10 contextual richness (28.7% of the variance) and entropy measures (37% of the variance). However, it has a limited effect on CC measures and is not significant for neighborhood density metrics. Figure 5 illustrates the marginal effects of the interaction of the two terms over TC_10. We can interpret this plot as follows: words with low specificity scores (red line) have lower context richness (TC), but within this group, the more words are concrete, the more they tend to have higher TC scores. However, this effect is reversed for highly specific words (blue line): TC scores tend to decrease for more concrete words.

A similar outcome is observed for the entropy measure (Figure 6). Generic words, both concrete and abstract, have a high entropy (pink line), meaning that these words are little expected given the context words. Conversely, specific words (green line) have a low entropy value, with abstract-specific words having lower entropy than concrete-specific words, meaning that abstract words are more predictable from context than concrete words.

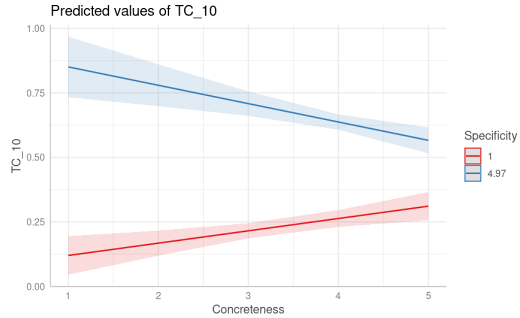


Figure 5: Interaction plot showing the relationship between Concreteness and TC_10 for different levels of Specificity (see also Appendix B).

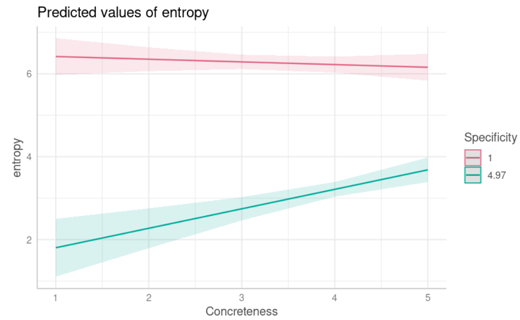


Figure 6: Interaction plot showing the relationship between Concreteness and entropy for different levels of Specificity.

The interaction models reveal a scenario that diverges from previous works: contextual variability does not depend on the dichotomy concrete-abstract, but more on the specificity of the word itself. Surprisingly, abstract-specific words like ‘bankruptcy’ have lower contextual variability than concrete-specific words like ‘hamburger’; that is, **abstract and specific words occur in a more**

limited and predictable number of selected contexts.

4.2 General observations

Comparing the linear models for the two spaces, the heatmaps in Figure 2 show that regression models are similar for neighborhood density (top of the heatmaps). This suggests that the two distributional spaces, while relying on different co-occurrence patterns, tend to build similar word representations. However, coefficients differ for context richness. High R^2 values are obtained considering the average cosine similarity between the target word and its context (TC) for the ITAw10 space in both Specificity and Interaction models, and average context-context (CC) similarity explains part of the variance in the Interaction model. Interestingly, ITAw2 shows an opposite trend: TC scores are not significant (Specificity and Interaction models), and a small variance is explained for CC values by the Specificity model. This outcome seems to confirm that a 2-word window is too small to extract useful distributional information. Overall, the analyses suggest that distributional measures are helpful for investigating cognitive assumptions, but the choice of the model could influence the final outcome.

We also run correlations across contextual variability measures in order to see how they overlap and complement each other. The main outcome is that TC_10 and entropy are strongly negatively correlated (Spearman's ρ = -0.713, $p < 0.001$), but only for ITAw10 space. As observed in the "Specificity effects" section, they represent the same distributional signature of a word but from a different perspective. Moreover, entropy negatively correlates with neighborhood density scores for both spaces. For instance, the correlation between entropy and TN_50 is ρ = -0.513 (ITAw10) and ρ = -0.472 (ITAw10), $p < 0.001$. In contrast, we see low or no correlations between neighborhood density and context richness measures. Correlation matrices are reported in Appendix C.

To conclude, while neighborhood density measures capture some information related to both Concreteness and Specificity, entropy and TC_10 are the best contextual variability metrics associated with Specificity. It is worth noticing that TC_10 was the best measure reported by Schulte im Walde and Frassinelli (2022), but for predicting the concreteness of a word in a pair.

5 Discussion and Conclusion

These analyses provide an enriched view of the relationship between abstraction and contextual variability compared to previous research. In particular, by adding a neglected aspect of abstraction, namely categorical Specificity, we observed that the difference in contextual variability is actually more dependent on Specificity than on Concreteness. These analyses provide an enriched view of the relationship between abstraction and contextual variability compared to previous research. In particular, by adding a neglected aspect of abstraction, namely categorical Specificity, we observed that the difference in contextual variability is actually more dependent on Specificity than on Concreteness. In particular: similar and targeted contexts occur with specific words, while generic words (both abstract and concrete) are associated with more extensive and heterogeneous contexts. To answer our initial research questions, therefore: concreteness explains part of the variation in contextual variability of nouns, but more variation is explained by specificity and by the interaction between the two variables.

Three key points that the current study makes: First, it revises various terminologies related to contextual variability. Second, it is the first study to directly explore contextual variability using the relationship between specificity and concreteness operationalized through human-generated ratings. Finally, it is the first study to conduct this analysis within the context of the Italian language. The outcomes hereby reported corroborate Bolognesi et al. (2020)'s argument: Categorical abstraction (specificity) is a variable that is deeply affected by language rather than by perceptual information, and therefore it has a stronger relationship with how words are used in context (contextual variability). Conversely, concreteness is less shaped by the patterns of linguistic occurrences, and arguably it is more deeply affected by perceptual experience.

Future investigations could focus on fine-grained analyses of different types of nouns, as well as on adjectives and verbs. Co-occurrence patterns differ across part-of-speech, but given the limited number of verbs (less than 60), we preferred to focus on nouns only. The present study opens the way to a new line of research in cognitive and computational linguistics and provides a promising different perspective on the analysis of concepts at different levels of abstraction.

Acknowledgements

Funded by the European Union (GRANT AGREEMENT: ERC-2021-STG-101039777). Views and opinions expressed are however those of the author(s) 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. We thank the reviewers for their useful suggestions.

References

Marco Baroni, Silvia Bernardini, Federica Comastri, Lorenzo Piccioni, Alessandra Volpi, Guy Aston, and Marco Mazzoleni. 2004. Introducing the la repubblica corpus: A large, annotated, tei (xml)-compliant corpus of newspaper italian. In *LREC*.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web: a collection of very large linguistically processed webcrawled corpora. *Language resources and evaluation*, 43:209–226.

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.

Georgiana Dinu, Nghia The Pham, and Marco Baroni. 2013. DISSECT - DIStributional SEmantics composition toolkit. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 31–36, Sofia, Bulgaria. Association for Computational Linguistics.

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.

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.

Thomas K Landauer and Susan T Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211.

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.

Jordan J Louviere, Terry N Flynn, and Anthony Alfred John Marley. 2015. *Best-worst scaling: Theory, methods and applications*. Cambridge University Press.

George A Miller and Christiane Fellbaum. 1991. Semantic networks of english. *Cognition*, 41(1-3):197–229.

Maria Montefinese, Ettore Ambrosini, Beth Fairfield, and Nicola Mammarella. 2014. The adaptation of the affective norms for english words (anew) for italian. *Behavior research methods*, 46:887–903.

Allan Paivio. 1991. Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 45(3):255.

Penny M Pexman, Ian S Hargreaves, Paul D Siakaluk, Glen E Bodner, and Jamie Pope. 2008. There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, 15:161–167.

Gabriel Recchia and Michael N Jones. 2012. The semantic richness of abstract concepts. *Front. Hum. Neurosci.*, 6:315.

Enrico Santus, Alessandro Lenci, Qin Lu, and Sabine Schulte Im Walde. 2014. Chasing hypernyms in vector spaces with entropy. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 38–42.

Paula J Schwanenflugel. 2013. Why are abstract concepts hard to understand? In *The psychology of word meanings*, pages 235–262. Psychology Press.

Paula J Schwanenflugel and Edward J Shoben. 1983. Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1):82.

Claude E Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Vered Shwartz, Enrico Santus, and Dominik Schlechtweg. 2017. Hypernyms under siege: Linguistically-motivated artillery for hypernymy detection. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 65–75, Valencia, Spain. Association for Computational Linguistics.

Gabriella Vigliocco, Stavroula-Thaleia Kousta, Pasquale Anthony Della Rosa, David P Vinson, Marco Tettamanti, Joseph T Devlin, and Stefano F Cappa. 2014. The neural representation of abstract words: the role of emotion. *Cerebral Cortex*, 24(7):1767–1777.

Sabine Schulte im Walde and Diego Frassinelli. 2022. Distributional measures of semantic abstraction. *Frontiers in artificial intelligence*, 4:206.

A Contextual Variability Measures

Measures of neighborhood density:

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

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

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

$$NN(t) = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k} similarity(i,j) \quad (3)$$

where $i \neq j$

Measures of context richness:

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

$$TC(t) = \frac{1}{k}\sum_{c=1}^{k} PPMI(t, c_i) \quad (4)$$

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

$$CC(t) = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k} similarity(i,j) \quad (5)$$

where $i \neq j$

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

$$DCR(t) = \frac{1}{k}\sum_{i=1}^{k} PPMI(t,i) \quad (6)$$

B Interaction plot

The plot reported in Figure 5 offers a graphical representation of the interaction (or relationship) between two continuous predictors, namely Concreteness and Specificity. In detail, we displayed the fitted values of the dependent variable (TC_10) on the y-axis and the values of the first factor (Concreteness) on the x-axis. The second factor (Specificity) is represented through lines on the chart – each possible value of the second factor gets its own line. As representative values of Specificity, we arbitrarily chose to plot only the two extreme values (1, 4.49 of the Specificity predictor. However, we could have plotted more values of Specificity (see Figure 7).

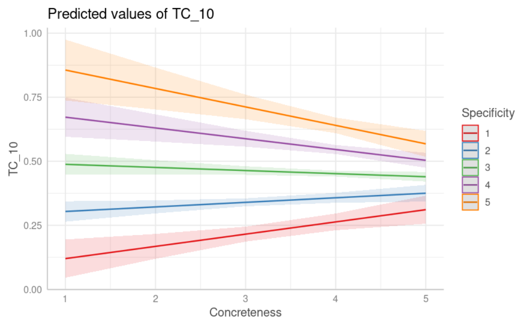


Figure 7: Interaction plot showing the relationship between Concreteness and TC0 for five different levels of Specificity.

C Correlations Between Measures

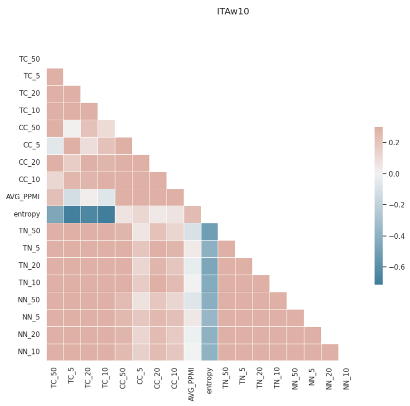


Figure 8: Spearman's ρ correlations among contextual variability measures for ITAw10.

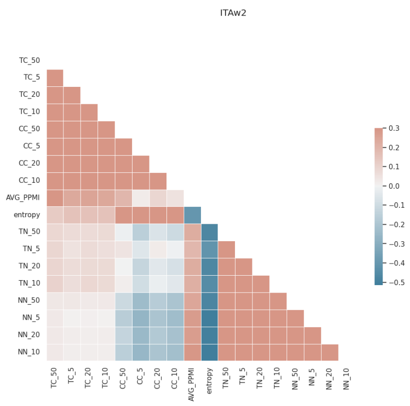


Figure 9: Spearman's ρ correlations among contextual variability measures for Itaw2.