

Contextual Characteristics of Concrete and Abstract Words

Diego Frassinelli, Daniela Naumann, Jason Utt, and Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart

[frassinelli|naumanda|uttjn|schulte]@ims.uni-stuttgart.de

Abstract

In this work we investigate commonalities and differences between the semantic representations of concrete and abstract words using human judgments and distributional semantics. We tackle the following questions: a) Does distributional similarity imply similarity in concreteness vs. abstractness? b) How do concrete and abstract context words co-occur with concrete and abstract words? c) Are our contextual models in line with existing theories of meaning representation? Our studies show that both distributionally similar words as well as distributionally co-occurring words come from the same range of concreteness vs. abstractness scores, partly challenging existing theories of semantic representation.

1 Introduction

The literature on conceptual representation has extensively debated about the nature of concrete concepts, but much less has been said about abstract concepts and about the similarities and differences between these two classes (Murphy, 2002). Multiple studies support the hypothesis that concrete concepts are directly grounded in the sensory-motor system, while abstract concepts are mapped to concrete concepts in order to be processed (Barsalou and Wiemer-Hastings, 2005; Hill et al., 2014; Pecher et al., 2011). Distributional semantics represents a very powerful approach to investigate word meaning in a data-driven fashion: the Distributional Hypothesis states that we can infer the meaning of a word by looking at the linguistic contexts it co-occurs with (Harris, 1954; Firth, 1957; Turney and Pantel, 2010). The resulting word representation, as a vector of co-occurrences of a word with the surrounding contexts, has been shown to be cognitively plausible (Miller and Charles, 1991; Lenci, 2008).

The aim of this work is to quantitatively investigate similarities and differences between concrete and abstract words by analysing the concreteness vs. abstractness of their respective linguistic contexts. Based on the literature, both concrete and abstract words should primarily co-occur with concrete words (i.e., their core semantic representation should mainly incorporate concrete words).

After describing the materials used, we will report three studies where: 1) we investigate the concreteness vs. abstractness of distributionally similar words, 2) we analyse the concreteness nature of co-occurring context words at type level, and 3) at token level.

2 Materials

For our studies, we selected nouns from the Brysbaert et al. (2014) collection of concreteness ratings for 40,000 English words. In this collection, each word was evaluated by at least 25 participants on a scale from 1 (abstract) to 5 (concrete). Given that participants were not aware of the part-of-speech of the word they were rating, we automatically assigned each word its more frequently occurring POS in the corpus. We focused our analyses on nouns because they are usually easier to classify according to their concreteness compared to adjectives and verbs. In total we had 9,241 nouns covered in an extensive selection of behavioural measures, such as valency scores (Warriner et al., 2013) and reaction times (Balota et al., 2007) which we aim to include in further analyses.

We used the selected nouns both as targets and as context words, and created a symmetric noun–noun co-occurrence matrix relying on a ± 20 word window in the ENCOW14A corpus, a collection of 16-billion English tokens extracted from the web (Schäfer, 2015). In this way, each co-occurrence is represented by a score (counts or positive LMI, cf. Evert (2004)). In addition, we ensure information about the concreteness scores for both the targets and the context words.

All the statistical analyses reported in this paper use linear mixed effects models (LME, Baayen et al. (2008)) with centered continuous predictors, implementing a maximal random effects structure as suggested by Barr et al. (2013).

3 Study 1: Investigating concreteness in distributionally similar words

In this study we investigate if distributionally similar words are also similar in their concreteness vs. abstractness scores. After computing the cosine similarity between each pair of target words using positive LMI transformation, we determined the nearest neighbours (NNs) for each target. Table 1 reports the top 8 NNs of the concrete word “lemon” (concreteness: 5.00) and the abstract word “belief” (concreteness: 1.19). The NNs are ordered by cosine similarity. The distributionally most similar words of the very concrete word “lemon” are, on average, also very concrete (6 out of 8 words have concreteness scores > 4). The two outliers “zest” and “concentrate” are ambiguous words with extremely different concreteness scores (e.g., “zest” means both “enthusiasm” and “skin”). In contrast, 4 of the 8 NNs of “belief” are very abstract words with concreteness scores < 2 . Its remaining NNs have a mid-range concreteness value and refer to specific groups of people associated by similar beliefs (e.g., “sect”).

Rank	NN	Similarity	Concreteness	Rank	NN	Similarity	Concreteness
-	lemon	-	5.00	-	belief	-	1.19
1	zest	0.831	2.27	1	spirituality	0.700	1.07
2	pineapple	0.756	4.94	2	atheist	0.642	2.93
3	cranberry	0.680	4.96	3	morality	0.634	1.47
4	ginger	0.617	4.92	4	superstition	0.593	2.07
5	grapefruit	0.601	4.96	5	sect	0.583	3.52
6	garnish	0.596	4.11	6	faith	0.561	1.63
7	concentrate	0.585	2.48	7	believer	0.546	2.73
8	orange	0.585	4.66	8	moral	0.539	1.69

Table 1: Top 8 NNs and their concreteness scores for the targets “lemon” and “belief”, sorted by cosine.

In a second step, we selected the top 2 to 16 NNs and averaged over their concreteness scores. Figure 1 presents these averages aggregated by the concreteness score of the corresponding targets (9 bins spanning .5 changes in score). The plot clearly shows that while increasing the concreteness score of the target (different lines), also the average concreteness ratings of the NNs increase (y-axis). An LME analysis indicates that this increase in the ratings associated with the increase of the target’s concreteness is statistically significant ($\beta_{concretenessTarget} = 0.22$, $p < .001$). On average, there are no significant differences between the different ranks ($p = 0.94$); however, there is a significant reduction in the ratings while increasing the number of NNs of highly concrete targets ($\beta_{concretenessTarget:rank} = -0.001$, $p < .001$).

Finally, we computed the neighbourhood density of each target (Sagi et al., 2009). Higher density scores indicate higher similarity between the vectors of the NNs and the vector of the corresponding target. Table 2 reports means and standard deviations of the neighbour density for each target concreteness bin. The right-most column reports the regression estimates of the pairwise comparison of each bin with its predecessor and the adjusted p-values. The analysis shows a higher neighbourhood density for the more extreme scores, both for concrete and abstract targets.

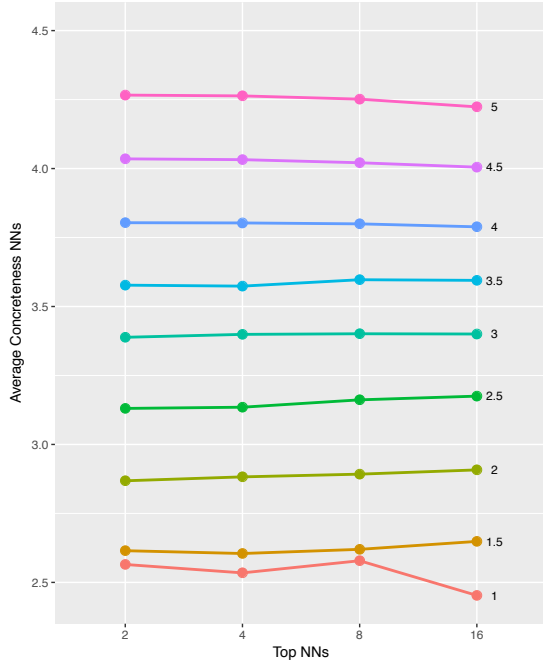


Figure 1: Study 1 - Average concreteness scores of the top 2-16 NNs grouped by the concreteness of the targets. The connection lines are not meaningful, but useful for visualisation purposes.

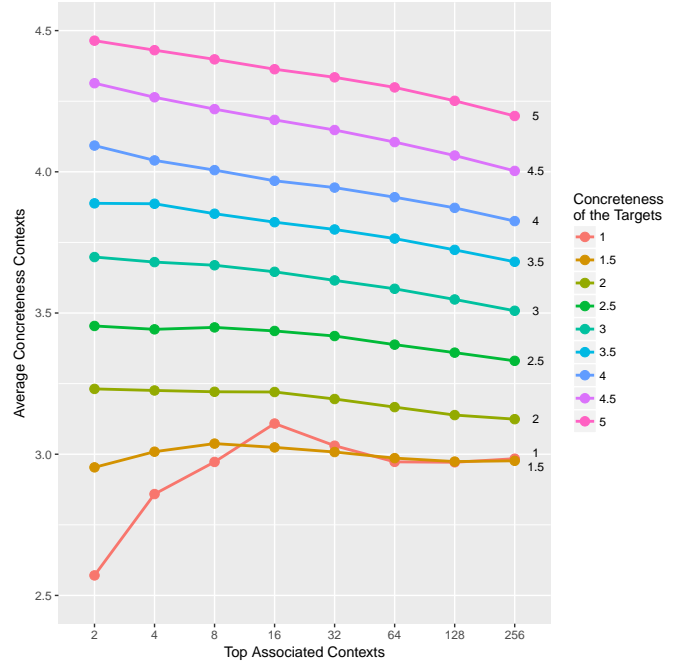


Figure 2: Study 2 - The five lines indicate the average concreteness scores of context words while decreasing their association with the target (2 to 256 co-occurring contexts words).

Concreteness	Density	SD	Estimate
1.0	0.468	0.083	
1.5	0.400	0.129	$\beta = -0.067$ ***
2.0	0.388	0.139	$\beta = -0.012$ ***
2.5	0.383	0.141	$\beta = -0.005$ *
3.0	0.393	0.144	$\beta = +0.009$ ***
3.5	0.391	0.147	$\beta = -0.002$ n.s.
4.0	0.402	0.147	$\beta = +0.012$ ***
4.5	0.415	0.146	$\beta = +0.013$ ***
5.0	0.434	0.145	$\beta = +0.019$ ***

Table 2: Average neighbourhood density and standard deviation (SD) for targets grouped in 9 bins according to their concreteness score. The Estimate column reports regression estimates and adjusted p-values comparing each bin with the previous one (e.g., 1.0 vs. 1.5).

To summarise, the main outcomes of this first study are: 1) distributionally similar words have a similar range of concreteness scores; 2) words with extreme concreteness scores have high neighbourhood density and are, on average, more similar to each other than mid-range words.

4 Study 2: Investigating the concreteness of context words at type level

In this study we shift our attention to the nature of the co-occurring context words. According to the Distributional Hypothesis, co-occurring words are essential elements for the definition of a word's meaning. For this reason, we investigate the concreteness patterns emerging from context words, in order to better understand the distributions of concrete and abstract words.

We grouped the targets into nine bins according to their concreteness ratings and we averaged the concreteness ratings of their first 2 to 256¹ most associated co-occurring context words (positive LMI scores). Figure 2 displays the outcome of this aggregation. An LME analysis indicates that the increase in the concreteness of the target corresponds to a significant increase in the average concreteness of its context words ($\beta_{concretenessTarget}=.189$, $p<.001$). The negative slopes of the lines in the plot indicate that less strongly associated context words are also less concrete ($\beta_{frequencyContext} = -.030$, $p<.001$). This pattern is more pronounced for the contexts of concrete targets ($\beta_{concretenessTarget:frequencyContext} = -.005$, $p<.001$).

Table 3 reports the 8 most frequent contexts of the words “lemon” and “belief”, and the context concreteness scores, sorted by LMI association. The 8 strongest co-occurring words of “lemon” are all extremely concrete (> 4.2), except for “zest” (see Study 1). Concreteness scores of the co-occurring words of “belief” are medium-low, with the exception of “people”.

Rank	Context	LMI	Concreteness	Rank	Context	LMI	Concreteness
-	lemon	-	5.00	-	belief	-	1.19
1	juice	47483.09	4.89	1	religion	154436.20	1.71
2	cup	27489.34	5	2	faith	85432.29	1.63
3	orange	21673.45	4.66	3	people	70812.64	4.82
4	sugar	15692.91	4.87	4	system	51274.14	2.94
5	dip	14138.21	4.22	5	freedom	31881.66	2.34
6	teaspoon	13608.63	4.76	6	practice	31808.87	2.52
7	zest	12910.50	2.27	7	value	30652.17	1.62
8	fruit	11633.53	4.81	8	attitude	30372.21	1.97

Table 3: 10 most associated contexts for the targets “belief” and “lemon” sorted by their positive LMI.

Overall, this second study highlights that target and context words share similar concreteness scores. Moreover, the concreteness of the contexts decreases with decreasing LMI association; and, in general, concrete words have more variability in the concreteness of their contexts compared to abstract words.

5 Study 3: Investigating the concreteness of context words at token level

In the previous study we only treated contexts as word types sorted by association score, but we did not completely exploit the informativeness of their co-occurrence strength. In order to analyse the distribution of concrete and abstract context words at token level, we now represent each target as a 9-dimensional concreteness vector, having one dimension per concreteness rating. Each dimension is the sum of frequencies of each context word having a specific concreteness rating, normalised by the total number of context words with the same score. In order to have all the values in the range 0-1, we normalised the scores in each cell again by the total amount of context words that each target has. For example, Table 4 reports the percentage of contexts with concreteness scores from 1 to 5 for “lemon” and “belief”. “Lemon” has a very high proportion of contexts with a concreteness of 5.0 (44%); while “belief” has 47% of its contexts with a concreteness score of 1 and 1.5.

	1	1.5	2	2.5	3	3.5	4	4.5	5
lemon	0.08	0.03	0.04	0.05	0.04	0.06	0.10	0.16	0.44
belief	0.18	0.29	0.16	0.10	0.08	0.07	0.04	0.04	0.04

Table 4: Percentage of tokens in each context bin for the words “lemon” and “belief”.

¹As discussed in Polajnar and Clark (2014), MI models achieve their best performance in semantic similarity tasks with vectors of 240 dimensions.

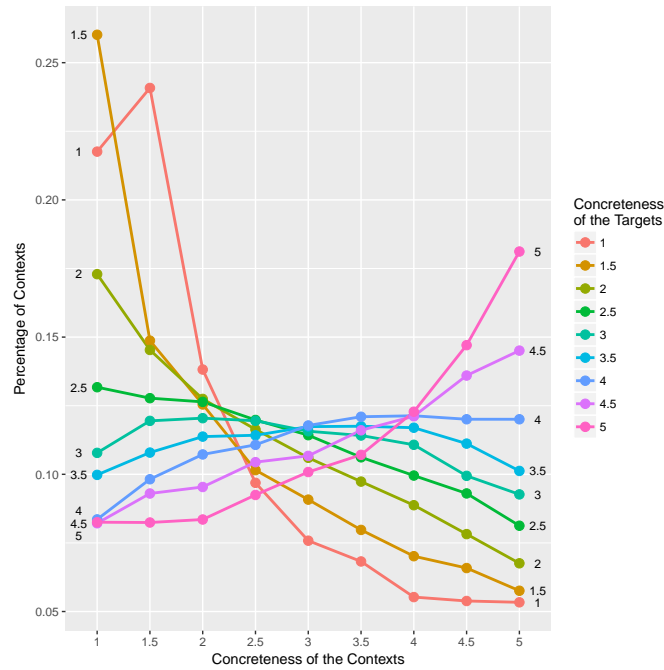


Figure 3: Study 3 - The five lines show the percentage of context words with a certain concreteness score averaged by the concreteness of their target.

Figure 3 shows an extremely clear picture. It reports the average number of contexts with a specific concreteness score (x-axis) grouped by the concreteness of the targets they co-occur with (different lines). Very concrete targets (e.g., pink and violet lines) have a clear preference for very concrete contexts (positive slope); on the other hand, very abstract words (red and yellow lines) show a clear preference for very abstract contexts (negative slope). The peak of the lines is steeper for abstract words than for concrete ones: compared to abstract words, concrete words seem to co-occur with words with different concreteness scores, still showing a clear preference for very concrete contexts.

Overall, Study 3 supports the evidence from the previous two studies: also at the token level, concrete words co-occur more frequently with concrete words, and abstract words co-occur primarily with abstract words.

6 Conclusion

Overall, the three studies show consistent results. Concrete words tend to co-occur with other concrete words, while abstract words tend to co-occur with abstract words. Moreover, concrete words seem to have more variable contexts in terms of their concreteness scores, compared to abstract words that seem to have a strong preference for abstract contexts with very low concreteness scores.

Our insights regarding concrete words are fully aligned with multiple studies in the literature (Barsalou and Wiemer-Hastings, 2005; Hill et al., 2014; Pecher et al., 2011). On the other hand, they seem to disagree with the grounding hypothesis for abstract words: in our studies, abstract words do not share the same context as concrete words. The importance of this research is threefold: it depicts a very consistent picture in the behaviour of concreteness measures from a distributional perspective; it also indicates some limitations of the behavioural measures adopted (e.g., average concreteness score for polysemous words); and it does not align with the existing psycholinguistic literature and thus provides a promising different perspective into the analysis of concrete and abstract concepts.

Acknowledgments

The research was supported by the DFG Collaborative Research Centre SFB 732 (Diego Frassinelli, Sabine Schulte im Walde, Jason Utt), and the DFG Heisenberg Fellowship SCHU-2580/1 (Sabine Schulte im Walde). We also thank the two anonymous reviewers for their comments.

References

- Baayen, R. H., D. J. Davidson, and D. M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4), 390–412.
- Balota, D. A., M. J. Yap, M. J. Cortese, K. A. Hutchison, B. Kessler, B. Loftis, J. H. Neely, D. L. Nelson, G. B. Simpson, and R. Treiman (2007). The English Lexicon Project. *Behavior Research Methods* 39(3), 445–459.
- Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 1–51.
- Barsalou, L. W. and K. Wiemer-Hastings (2005). Situating abstract concepts. In D. Pecher and R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking*, Chapter 7, pp. 129–163. New York: Cambridge University Press.
- Brysbaert, M., A. B. Warriner, and V. Kuperman (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46(3), 904–11.
- Evert, S. (2004). The statistical analysis of morphosyntactic distributions. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pp. 1539–1542.
- Firth, J. R. (1957). *Papers in Linguistics 1934-51*. London, UK: Longmans.
- Harris, Z. (1954). Distributional structure. *Word* 10(23), 146–162.
- Hill, F., A. Korhonen, and C. Bentz (2014). A quantitative empirical analysis of the abstract/concrete distinction. *Cognitive Science* 38(1), 162–177.
- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Rivista di Linguistica* 20(1), 1–31.
- Miller, G. A. and W. G. Charles (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6(1), 1–28.
- Murphy, G. L. (2002). *The Big Book of Concepts*. London: MIT Press.
- Pecher, D., I. Boot, and S. Van Dantzig (2011). Abstract concepts. Sensory-motor grounding, metaphors, and beyond. *Psychology of Learning and Motivation - Advances in Research and Theory* 54, 217–248.
- Polajnar, T. and S. Clark (2014). Improving distributional semantic vectors through context selection and normalisation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 230–238.
- Sagi, E., S. Kaufmann, and B. Clark (2009). Semantic density analysis: Comparing word meaning across time and phonetic space. In *Proceedings of the EACL Workshop on GEMS: GEometrical Models of Natural Language Semantics*, pp. 104–111.
- Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora*, pp. 28–34.

Turney, P. D. and P. Pantel (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37, 141–188.

Warriner, A. B., V. Kuperman, and M. Brysbaert (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45(4), 1191–1207.