

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA



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

Giulia Rambelli¹ Marianna M. Bolognesi¹

¹University of Bologna

Introduction

Cognitive and corpus-based studies observed that **concrete** words (*fork*) occur in a few but very similar linguistic contexts, while **abstract** words (*joy*) tend to be used in a wide variety of linguistic contexts.

However, no previous work has been carried out on words annotated for

Methods

Contextual Variability (CV) Measures

Compute the differences between the contexts of occurrence [2].

1. NEIGHBORHOOD DENSITY

TN: avg similarity between t and its k-nearest neighbors
 NN: avg similarity between the k-nearest neighbors of t

Concreteness and Specificity [1].

Concreteness happiness \rightarrow banana



Contributions

- Hypothesis Differences in the contextual associations are dependent on Specificity rather than Concreteness
- Novelty

Explore variations in Contextual Variability in English and Italian

2. CONTEXT RICHNESS

- **TC**: avg similarity between *t* and its k-top contexts
- **CC**: avg similarity between the k-top contexts of t
- **DCR**: mean of the PPMI scores of the k-top contexts of *t*.
- 3. ENTROPY

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

Data 676 English nouns from ANEW
DSM Context from ukWaC (window=10), embeddings by word2vec
Comparison with Italian data [3]

Analysis

Summary of linear models

Regression analyses investigate how CV measures are explained by the level of Concreteness and Specificity of words.

		Ν		R-squared
TN_5	4.74%***	2.83%***	4.54%	- 0.30
TN 10	F 650/***	2 200/ ***	5 17%	

1 CONCRETENESS EFFECTS

- NN and TC measures are explained by Concreteness (≠Italian).
- High correlation between Concreteness and TC10 (.44) and NN5 (.42)

3 INTERACTION EFFECTS

 TC and entropy show an interaction between Concreteness and Specificity.

Pre	edicted	value	s of en	tropy		

111_10	5.05%	5.29%	5.47%	
TN_20	5.88%***	3.03%***	5.64%	0.25
TN_50	5.44%***	2.22%***	5.19%	- 0.25
NN_5	9.88%***	6.9%***	10.09%	
NN_10	12.46%***	8.71%***	12.77%	0.20
NN_20	13.58%***	7.56%***	13.46%	- 0.20
NN_50	13.83%***	6.45%***	13.57%	
UD TC_5	15.8%***	11.18%***	16.81%*	0.15
DTC_5 TC_10	17.06%***	13.3%***	18.77%*	- 0.15
TC_20	12.52%***	8.99%***	13.67%*	
TC_50	6.51%***	2.52%***	7.1%*	0.10
CC_5	3.0%***	2.53%***	3.76%*	- 0.10
CC_10	3.2%***	4.03%***	5.51%**	
CC_20	0.94%**	2.11%***	3.45%***	0.05
CC_50	0	0	0.42%*	- 0.05
DCR	0	0	1.49%**	
Н	0.13%	14.96%***	26.8%***	0.00
	CV~Conc	CV~Spec model	CV~Conc*Spec	- 0.00

Figure 1. Cells report Adjusted R^2 values and *p*-values. ... p < 0.1, *=p < .05, **=p < .01, and ***=p < .001.

2 SPECIFICITY EFFECTS

- Contexts vary depending on the Specificity of a word independently of its Concreteness.
- Negative correlation between Entropy and Specificity (ρ =-0.42).

pasta	food
spec: 4.2, conc: 4.7	spec: 1.5, conc: 4.8
dish(.6), sauce (.8)	eat (.6), find (.3)
bread (.6), rice (.6)	drink (.6), chain (.3)
food (.5), salad (.8)	fast (.2)

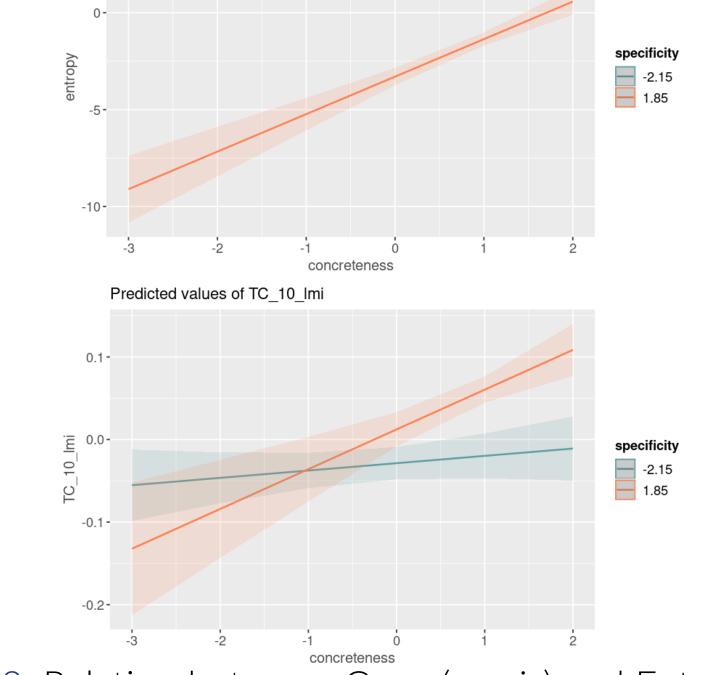


Figure 2. Relation between Conc (x-axis) and Entropy (top) / TC_10 (bottom) for different levels of Spec.

giulia.rambelli4@unibo.it

Additional Observations

Italian vs. English Correlations between Italian and English metrics.

Conc Spec TN5 NN5 TC5 CC5 H DCR 0.797 0.824 0.323 0.158 0.238 0.196 0.746 0.442

 Specific words have well-defined, similar contexts. Generic words, whether abstract or concrete, have broader contexts.

Conclusive Remarks

 Concreteness is more significant in explaining noun CV in English than in Italian.

Analysis of Contexts

- Concreteness of contexts is higher for more concrete nouns (ρ =.71)
- More specific words tend to have more concrete contexts (ρ =.62)
- Specificity of context words varies depending on the target's Concreteness (ρ =.42) or Specificity (ρ =.46)
- The interaction between Concreteness and Specificity accounts for a significant portion of the variation in the regression analyses.
- Entropy is cross-linguistically reliable, while metrics from DSM are more language-dependant.

References

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.

- [1] M. Bolognesi, C. Burgers, and T. Caselli. On abstraction: decoupling conceptual concreteness and categorical specificity. *Cognitive Processing*, 21(3):365–381, 2020.
- [2] S. Schulte im Walde and D. Frassinelli. Distributional measures of semantic abstraction. *Frontiers in artificial intelligence*, 4:206, 2022.
- [3] G. Rambelli and M. Bolognesi. Contextual variability depends on categorical specificity rather than conceptual concreteness: A distributional investigation on Italian data. In *Proceedings of IWCS* 2023, 2023.

LREC-COLING 2024, 22 May, Turin