

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 Concreteness and Specificity [1].

Concreteness
happiness → *banana*

Specificity
animal → *cat*

Contributions

- **Hypothesis** Differences in the contextual associations are dependent on Specificity rather than Concreteness
- **Novelty** Explore variations in Contextual Variability in English and Italian

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*
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**

$$H(w) = - \sum_c p(c|w) * \log_2(p(c|w))$$

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.

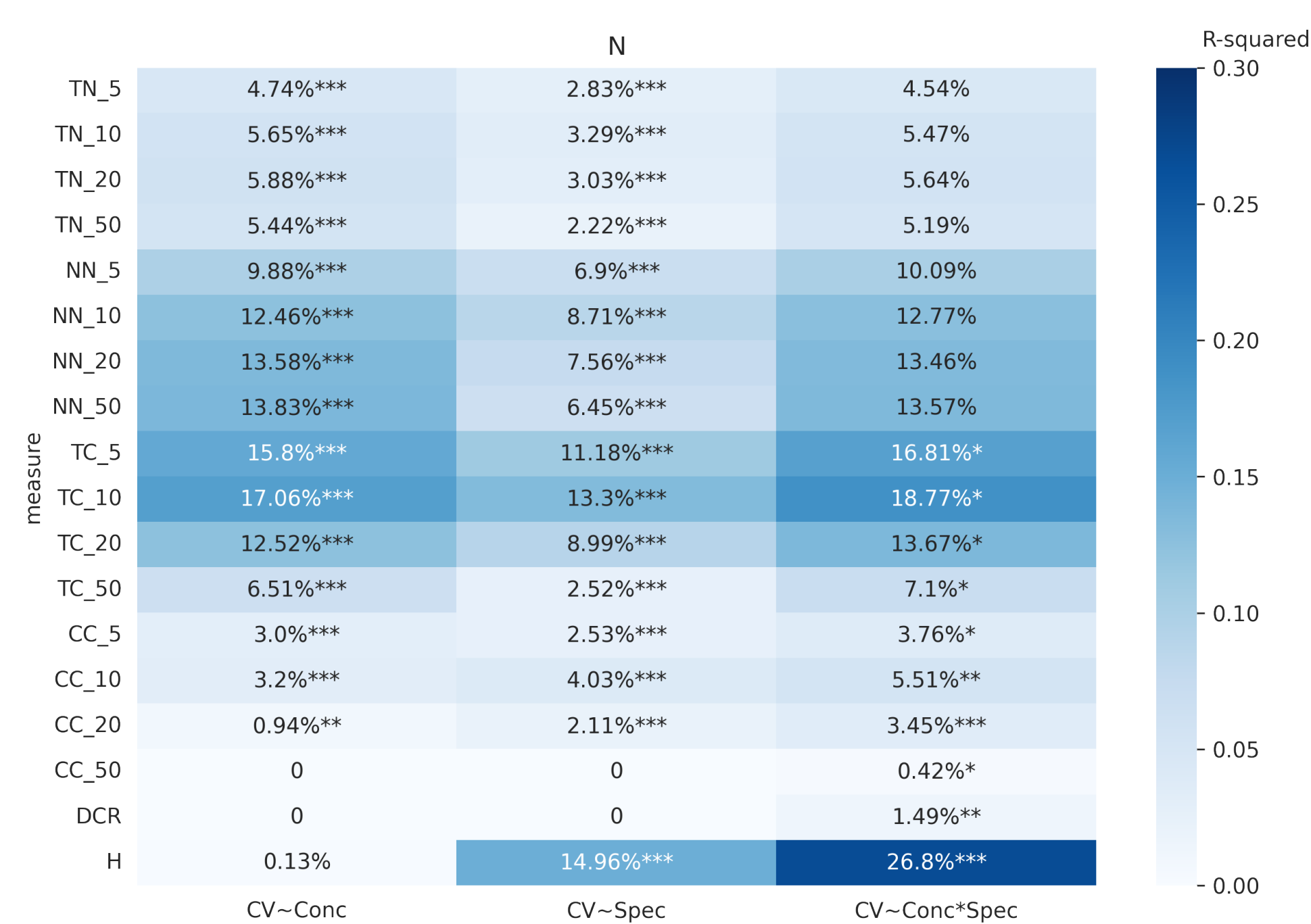


Figure 1. Cells report Adjusted R^2 values and p -values. $^{\circ}$ = $p < 0.1$, * = $p < .05$, ** = $p < .01$, and *** = $p < .001$.

1 CONCRETENESS EFFECTS

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

2 SPECIFICITY EFFECTS

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

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

3 INTERACTION EFFECTS

- TC and entropy show an interaction between Concreteness and Specificity.

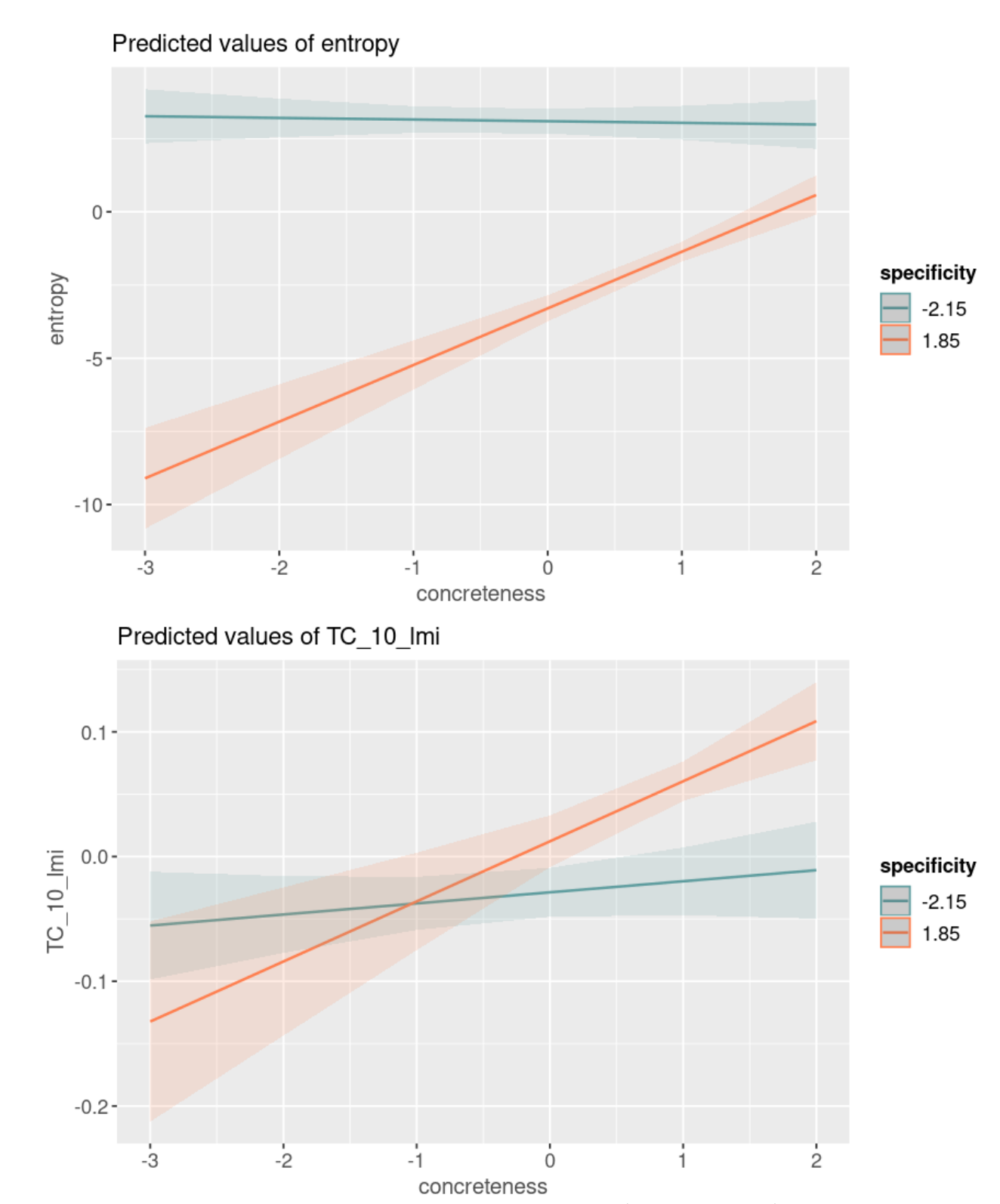


Figure 2. Relation between Conc (x-axis) and Entropy (top) / TC_10 (bottom) for different levels of Spec.

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

Analysis of Contexts

- Concreteness of contexts is higher for more concrete nouns ($\rho = .71$)
- More specific words tend to have more concrete contexts ($\rho = .62$)
- Specificity of context words varies depending on the target's Concreteness ($\rho = .42$) or Specificity ($\rho = .46$)

Conclusive Remarks

- **Specific words** have well-defined, similar contexts. **Generic words**, whether abstract or concrete, have broader contexts.
- Concreteness is more significant in explaining noun CV in English than in Italian.
- 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

- [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.