The role of meaning in the rivalry of *-ity* and *-ness*: evidence from distributional semantics

Martin Schäfer
post@martinschaefer.info

Düsseldorf, August 8, 2023

Introduction

The *-ity* and *-ness* affix rivalry:

- frequent and productive suffixes
- same core function
- vast majority of bases take either *-ity* or *-ness*, but doublets exist

(1) *-ity*

a. insular: insularity
b. eatable: eatability
c. sentimental: sentimentality

(2) *-ness*

a. red: redness
b. messy: messiness
c. pleasant: pleasantness

(3) *-ity* and *-ness*

a. aggressive: aggressivity/aggressiveness
b. opportune: opportunity/opportuneness
c. casual: casuality/casualness

Introduction ctd

1. What determines the choice between *-ity* and *-ness* for a given base word?
2. Are the two affixes synonyms?

- Why *insularity* and *redness* and not *redity* and *insularness*?
- Any systematic meaning differences between doublets like *aggressivity*/*aggressiveness*?

`Note: the study is restricted to adjectival bases!`

Background: bases

Constraints and patterns

- based on morphological make-up of the base (Lindsay, 2012)
- based on form features of the base (Arndt-Lappe, 2014)
- based on semantics of the base (Riddle, 1985)
- *able/-ible* $\rightarrow$ *-ity*; *-less* $\rightarrow$ *-ness*
- *-ile*: *sterile/vile*
- color words; meaning encoded in morphemes

Background: synonyms

- Standard view (Marchand, 1969): both form abstract substantives; "state, quality, condition of BASE"
- Non-synonym view (Riddle, 1985): "*-ness* tends to denote an embodied attribute or trait, while *-ity* tends to denote an abstract or concrete entity."

(4) a. "However, don't call this third-grader a picky eater. She's a selective one, a Feingold diet subscriber, whose *hyperactiveness* has decreased, her mother says, since she began the program four years ago."

b. "But to date there is no evidence that this type of dietary regime will have any effect on *hyperactivity* in children."

`Examples from Riddle; contra Riddle: Bauer, Lieber, and Plag (2013)`

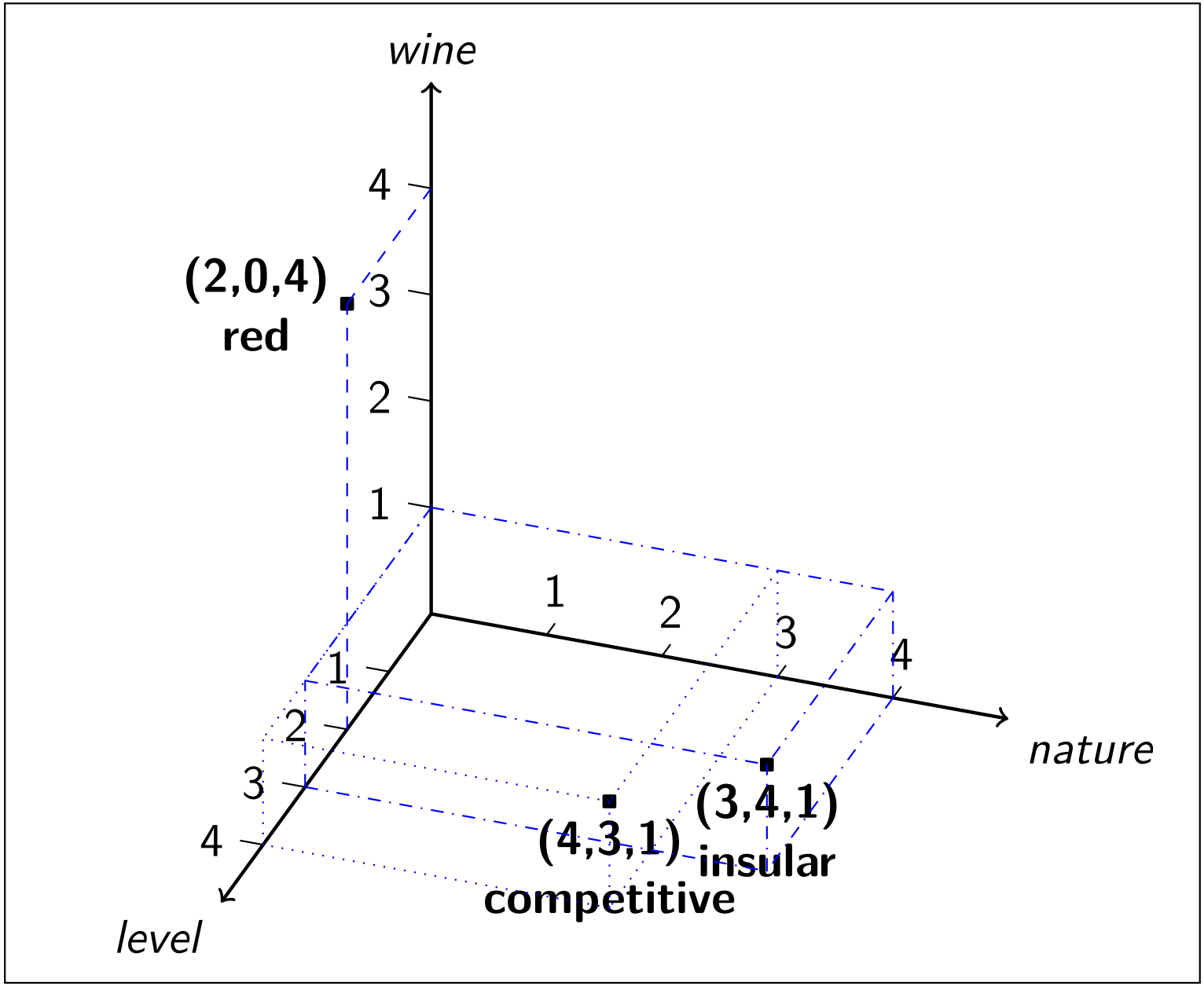
Distributional semantics

The distributional hypothesis:
Words with similar distributional properties have similar meanings.
Sahlgren (2006, p. 21)

target words	cooccurrences with *level*	*nature*	*wine*
competitive			
red			
insular			

target words	cooccurrences with *level*	*nature*	*wine*
competitive	4	3	1
red	2	0	4
insular	3	4	1

Mapping into geometrical space



Hypotheses

(1) Does base semantics drive affix selection?
  (1a) Clear difference between vectors of *-ity* bases and vectors of *-ness* bases
  (1b) Difference should obtain for bases with the same endings

(2) Are the two affixes synonyms?
  (2a) If *-ity*/*-ness* are synonyms, same shift in semantic space
  (2b) Doublets (such as *aggressivity*/*aggressiveness*) without systemantic semantic differences

Methods: material

- Pre-trained word embeddings: fastText vectors (Mikolov et al., 2017)

(1) base semantics: 1345 *-ity* and 1671 *-ness* pairs, doublets are excluded (*aggressive* → *aggressivity*/*aggressiveness*)
   - Subset of 198 *-ive* bases
   - 90 with *-ity* derivatives: *relative*
   - 108 with *-ness* derivatives: *distinctive*

(2) synonyms or not:
   - all derivatives of the non-doublets
   - 131 doublets

Methods: analysis

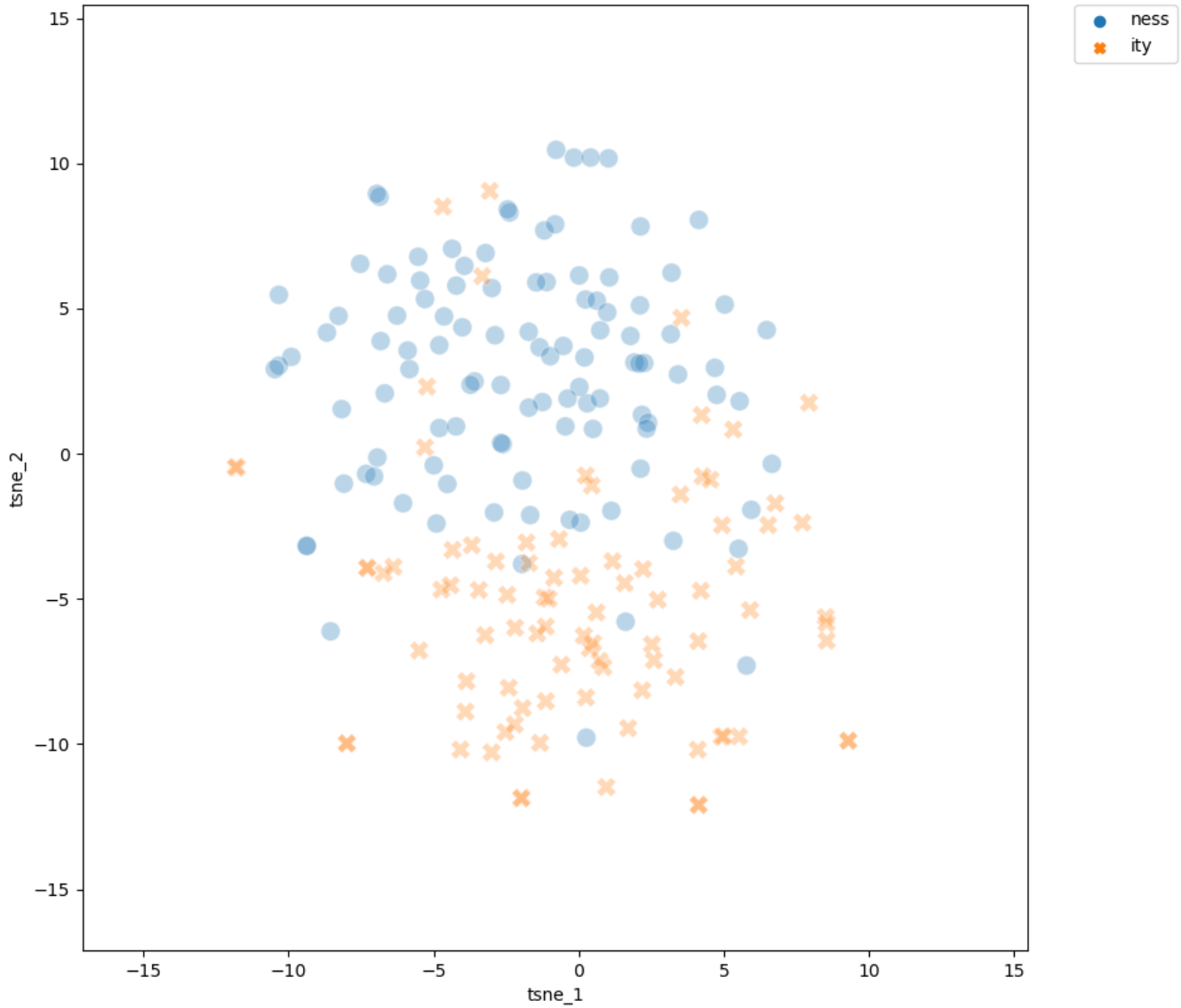
- Clustering with t-Distributed Stochastic Neighbor Embedding (t-SNE) (Maaten and Hinton 2008)
- Linear Discriminant Analysis (LDA) for statistical corroboration

For the t-SNE/LDA pipeline, cf. Shafaei-Bajestan et al. (2022)

Results 1a: all non-doublet bases

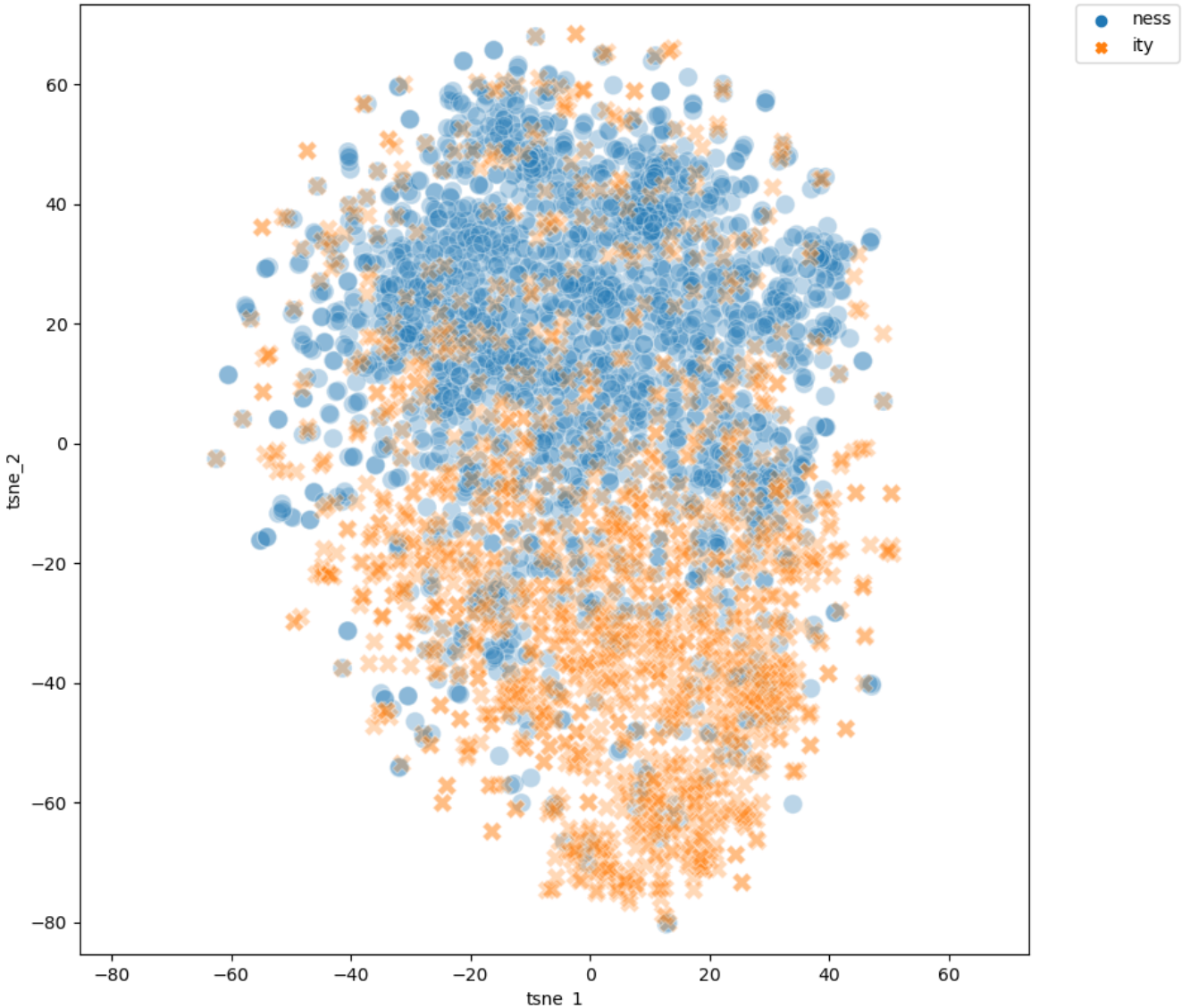
LDA: average weighted F1 score = 0.849 (0.017 std);
baseline classifer: 0.395)

Results 1b: *-ive* bases



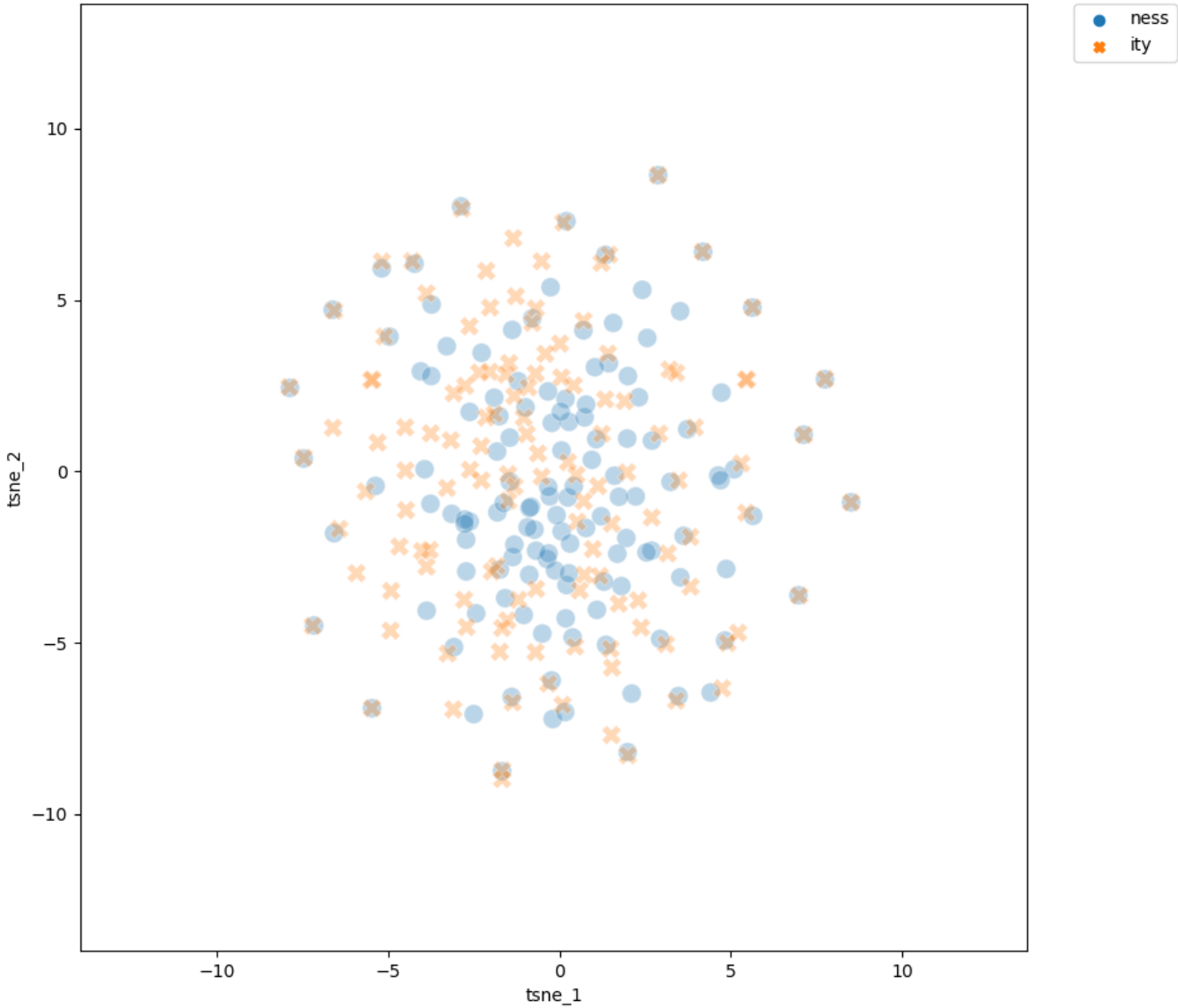
LDA: mean weighted F1 score: 0.744, std 0.098; 0.385 baseline classifier

Results 2.1



LDA: mean weighted F1 score = 0.859, std = 0.018; 0.385 baseline classifier

Results 2.2



LDA: mean weighted F1 score 0.583 (0.08 std); 0.333 baseline classifier

Conclusion

- Meaning of the bases is a major factor in affix selection:
  - Across all non-doublet bases
  - Even for all non-doublet *-ive* bases
- Affixes are synonyms
  - Affixation induces similar shifts
  - No systematic patterns in doublets
- Next steps
  - Zooming on the properties that are behind the distinct vector characteristics
  - Direct comparison to form-based models

Thank you!

References I

- Arndt-Lappe, Sabine (Nov. 2014). "Analogy in suffix rivalry: the case of English -ity and -ness". In: *English Language and Linguistics* 18 (03), pp. 497–548. ISSN: 1469-4379. DOI: 10.1017/S136067431400015X. URL: http://journals.cambridge.org/article_S136067431400015X.
- Bauer, Laurie, Rochelle Lieber, and Ingo Plag (2013). *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
- Lindsay, Mark (2012). "Rival suffixes: synonymy, competition, and the emergence of productivity". In: *Morphology and the architecture of grammar: Proceedings of the 8th International Morphology Meeting*. Ed. by Angela Ralli et al. Vol. 8. University of Patras. Patras, pp. 192–203.

References II

- Maaten, Laurens van der and Geoffrey Hinton (2008). "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9, pp. 2579–2605.
- Marchand, Hans (1969). *The Categories and Types of Present-Day English Word-Formation. A Synchronic-Diachronic Approach*. 2nd, completely revised and enlarged. München: C.H. Beck'sche Verlagsbuchhandlung.
- Mikolov, Tomás et al. (2017). "Advances in Pre-Training Distributed Word Representations". In: *CoRR* abs/1712.09405. URL: http://arxiv.org/abs/1712.09405.
- Riddle, Elizabeth M. (1985). "A historical perspective on the productivity of the suffixes -ness and -ity". In: *Historical Semantics— Historical Word-Formation*. Ed. by Jacek Fisiak. De Gruyter Mouton, pp. 435–462. DOI: doi:10.1515/9783110850178.435. URL: https://doi.org/10.1515/9783110850178.435.

References III

- Sahlgren, Magnus (2006). "The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces". PhD thesis. Stockholm: Stockholm University.
- Shafaei-Bajestan, Elnaz et al. (2022). *Semantic properties of English nominal pluralization: Insights from word embeddings*. DOI: 10.48550/ARXIV.2203.15424. URL: https://arxiv.org/abs/2203.15424.

Similarities within doublets

Considerable variation: minimum of 0.268 and a maximum of 0.867 (median = 0.639, mean = 0.614).

Table: Illustration of doublets across the distribution of cosine similarities within doublets. The two doublets closest to the respective values have been selected.

place within distribution	doublet
Min (0.2680)	*opportunity/opportuneness* *casuality/casualness*
1st Qu. 0.5423	*naturality/naturalness* *obliquity/obliqueness*
Mean 0.6137	*chastity/chasteness* *changeability/changeableness*
3rd 0.7211	*exhaustivity/exhaustiveness* *passivity/passiveness*
Max 0.8671	*impassivity/impassiveness* *inclusivity/inclusiveness*

Modeling the similarity

Table: Beta regression for cosine similarity between the doublets.
R-sq.(adj) = 0.14 Deviance explained = 16.5%

Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.369267	0.227343	6.023	1.71e-09
ityLogFreq	-0.164101	0.037082	-4.425	9.63e-06
nessLogFreq	-0.158859	0.054403	-2.920	0.00350
ityLogFreq:nessLogFreq	0.030802	0.009461	3.256	0.00113

Interaction plots