The role of semantics in the rivalry of *-ity* and *-ness*: Evidence from distributional models

Martin Schäfer
HHU Düsseldorf, post@martinschaefer.info

Keywords: Morphology, English, Affix rivalry, Distributional Semantics

The *-ity/-ness* affix rivalry (*rationality*/*happiness*) raises two core questions: (1) What determines the choice between *-ity* and *-ness* for a given base word (Lindsay, 2012)? (2) Are the two affixes synonyms? Arndt-Lappe (2014) shows that phonological, not morphological, properties of endings are decisive, but she does not consider potential semantic influences. Riddle (1985) argues that the two suffixes have different meanings, whereas Bauer et al. (2013) find no productive semantic difference.

We use distributional semantics to address both questions, hypothesizing that: (1a) If the semantics of the bases drives affix selection, a clear semantic difference between bases taking *-ity* and bases taking *-ness* is expected. (1b) This difference should even obtain for bases that end in the same suffix (e.g. *-ive*). (2a) If *-ity/-ness* are synonyms, the shift in semantic space induced by the two suffixes should be the same for both *-ity* and *-ness* derivatives. (2b) Doublets (such as *aggressivity/aggressiveness*) should show no systematic semantic difference.

To test these hypotheses, we used all pairs of adjectival bases and *-ity/-ness* derivates in the ukWaC corpus (Baroni et al., 2009) with pretrained fasttext vectors (Mikolov et al., 2017) (1546 *-ity*/1835 *-ness* pairs). Three subsets were formed: The first comprised all base-derivative pairs without doublets. The second consisted of all doublet-less pairs that contained bases ending in the suffix *-ive*, the only ending with an adequate number of both derivates (90 *-ity* and 108 *-ness*). The third subset contained all (and only) doublets. To illustrate, the base-derivative pairs *insular-insularity* and *red-redness* are part of the first subset, because neither derivative occurs in a doublet (the strings *redity* and *insularness* do not occur in the dataset). Pairs like *narrative-narrativity* and *distinct-distinctiveness*, where again the derivatives do not occur in doublets, form the second subset. The third subset contains only doublets, e.g. *aggressivity/aggressiveness* or *naturality/naturalness*. Since the bases of doublets cannot be distinctive, they are not considered in any of the analyses. Vectors are analysed with the t-SNE visualization method (van der Maaten & Hinton 2008, Arora et al. 2018), which we statistically corroborated by Linear Discriminant Analysis (LDA) (following Shafaei-Bajestan et al. 2022).

Figures (1a) through (2b) show the t-SNE visualizations. Blue circles represent the projections of the vectors of the bases with *-ness* derivates or the vectors of *-ness* derivates themselves, red crosses represent the corresponding *-ity* vectors.

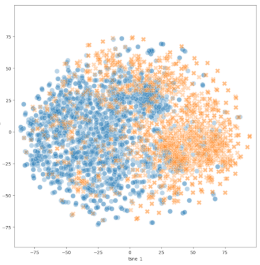

Fig. 1a: all bases (no doublets)

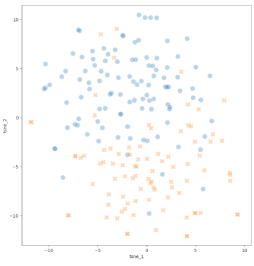

Fig. 1b: all -ive bases (no doublets)

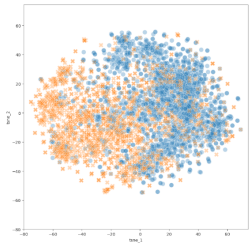

Fig. 2a all derivates (no doublets)

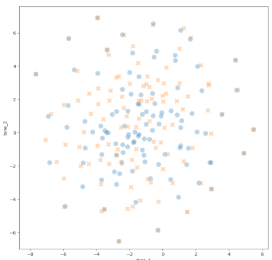

Fig. 2b: all derivate doublets

Figure 1a shows that the *-ness* bases cluster on the left of the plot, the *-ity* bases on the right, with individual and small clusters of outliers for both large clusters. Figure 1b shows similarly clear clustering for the *-ive* bases (absolute orientation of the projections is meaningless). The patterning of the data as shown in figures 1a and 1b thus supports hypotheses (1a) and (1b): the semantics of the base forms is closely associated with a preference for either *-ity* or *-ness*. Figure 2a shows that the derivates cluster in

a similar fashion as do their bases. This is in line with hypothesis (2a): the extent to which the derivates differ from their bases is similar for both affixes, as expected for synonyms. Regarding hypothesis (2b), figure 2b reveals no clustering, again in line with *-ity* and *-ness* being synonyms.

This study reveals the base semantics is a major factor in affix selection. Our results suggest that Arndt-Lappe's phonological effect of the word's endings emerges from the shared semantics of the respective bases.

References

Arndt-Lappe, Sabine. 2014. Analogy in suffix rivalry: the case of English -ity and -ness. *English Language and Linguistics* 18. 497–548. doi:10.1017/S136067431400015X. http://journals.cambridge.org/article_S136067431400015X.

Arora, Sanjeev, Wei Hu & Pravesh K Kothari. 2018. An analysis of the t-SNE algorithm for data visualization. In *Conference on learning theory*, 1455–1462. PMLR.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226. doi:10.1007/s10579-009-9081-4.

Bauer, Laurie, Rochelle Lieber & 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 Angela Ralli, Geert Booij, Sergio Scalise & Athanasios Karasimos (eds.), *Morphology and the architecture of grammar: Proceedings of the 8th international morphology meeting*, vol. 8, 192–203. Patras: University of Patras.

van der Maaten, Laurens & Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9. 2579–2605.

Mikolov, Tomás, Edouard Grave, Piotr Bojanowski, Christian Puhrsch & Armand Joulin. 2017. Advances in pre-training distributed word representations. *CoRR* abs/1712.09405. http://arxiv.org/abs/1712.09405.

Riddle, Elizabeth M. 1985. *A historical perspective on the productivity of the suffixes -ness and -ity* 435–462. De Gruyter Mouton. doi:doi:10.1515/9783110850178.435. https://doi.org/10.1515/9783110850178.435.

Shafaei-Bajestan, Elnaz, Masoumeh Moradipour-Tari, Peter Uhrig & R. Harald Baayen. 2022. Semantic properties of english nominal pluralization: Insights from word embeddings. doi:10.48550/ARXIV.2203.15424. https://arxiv.org/abs/2203.15424.