

Nominalizations with the suffixes *-ee* and *-ation* – A distributional semantic analysis

Viktoria SCHNEIDER and Ingo PLAG

Nominalizations with derivational suffixes can give rise to interpretations as eventualities, for example, processes or states, like *articulation*, *levitation*. Alternatively, nominalizations may also refer to participants in eventualities, like *employee*, *nominee*. The derivatives which denote such eventuality-related interpretations are clearly semantically related to the eventuality denoted by their base word (see, e.g., Barker 1998; Plag et al. 2018; Kawaletz 2021; Schneider 2022). With verbs as bases, such eventuality-related nominalizations are straightforward, as verbs ontologically denote eventualities themselves. With nouns, on the other hand, eventuality-related nominalizations are less straightforward due to their ontology, as they standardly refer to entities, not eventualities (see, e.g., Van Valin & LaPolla 2002; Haspelmath 2001; Szabó 2015). For nominal bases, more semantic decomposition of the base, or inferencing, is needed to identify the relevant eventuality that can be used for the interpretation of the nominalization (e.g., Schneider 2022). Examples of denominal eventuality-related nominalizations are *sedimentation*, *ozonation*, *biographee*, and *debtee*. The exact nature of the semantics of such denominal eventuality-related nominalizations is largely unclear, since the research on eventuality-related nominalizations focused almost entirely on those with verbal bases.

Based on results emerging from studies of eventuality-related nominalizations, two hypotheses can be formulated. First, deverbal derivatives and their verbal bases should be highly similar in their meaning as they operate on the same eventuality. Second, denominal derivatives and their nominal bases, in contrast, should be less semantically similar to each other, since the pertinent eventuality is not as easily accessible for the word formation process in the first place. These two hypotheses are investigated in the present study.

Distributional semantics has been shown to be a fruitful approach to test semantic similarities and dissimilarities of derivatives (see, e.g., Lapesa et al. 2018; Wauquier et al. 2018; Huyghe & Wauquier 2020). For the present study, to use such a distributional semantic approach, word vectors were computed using fastText (Bojanowski et al. 2016; Mikolov et al. 2018). Vectors of eventuality-related derivatives with the suffix *-ee* and *-ation* and their pertinent bases were then compared using

cosine similarity. As these vectors represent a word's semantics, these cosine similarities, in turn, represent the semantic similarities between words.

We used the cosine similarities to measure the strength of the semantic relation between base and derivative by word class. This strength of relation may, however, also be influenced by at least two other factors. First, there is the relative frequency of a base and its derivative. Higher relative frequency is said to lead to a better segmentability of the resulting derivative (see, e.g., Hay & Baayen 2003). This effect should lead to an overall higher similarity of base and derivative because the connection of the two words is more clearly identifiable. Second, the polysemy of the base word also influences the cosine similarity. Derivatives usually instantiate one specific reading of the base word. A multitude of readings of a base word leads to a semantic vector for that word that aggregates over all readings. The similarity between a highly polysemous base word and its derivative should therefore be smaller than between a less polysemous base word and its derivative.

A multivariate statistical procedure is needed to take all three things into account at the same time. As the cosine similarities in this study are distributed over an interval between 0 and 1, beta regression is the model of choice. Word class of the base, relative frequency, and polysemy of the base showed high correlation coefficients for the set of *-ation* bases and derivatives ($|r_{rrhoo}| > 0.5$). To address this potential collinearity problem, we devised a principal component analysis for these three variables. (see, e.g., Baayen 2008; Tomaschek et al. 2018). For the set of *-ee* derivatives and their bases, no high correlation coefficients were found. Thus, for *-ee*, the three variables entered the modelling process directly.

The results show significant differences in cosine similarity for denominal and deverbal derivatives and their bases. For derivatives with the suffix *-ee*, denominal derivatives and their nominal bases are significantly more similar to each other than deverbal derivatives and their verbal bases are. In contrast, for derivatives with the suffix *-ation*, deverbal derivatives and their verbal bases are more similar to each other than denominal derivatives and their nominal bases. Relative frequency decreases the cosine similarity of derivatives and bases for both suffixes. This is opposite to the prediction. Polysemy of the base decreases the cosine similarity of derivative and base, as expected, but only significantly so for nominalizations with the suffix *-ation*.

Our results show that eventuality-related nominalizations with the suffix *-ation* show the expected similarities of derivatives and bases: The deverbal derivatives and their verbal bases are more similar to each other than the denominal derivatives and

their nominal bases. The nominalizations in *-ee* did not show the expected pattern. We interpret this difference as emerging from a difference in the semantic output category of the suffixes. Derivatives with the suffix *-ee* create a participant reading, and such a reading is ontologically more closely related to nouns (see, e.g., Barker 1998; Plag 2004; Bauer et al. 2013; Plag et al. 2018; Schneider 2022). Derivatives with *-ation*, on the other hand, describe mostly processes (see, e.g., Bauer et al. 2013; Plag 2018) which are ontologically more related to verbs (see, e.g., Van Valin & LaPolla 2002; Haspelmath 2001; Szabó 2015). These findings show that not only the type of base influences the semantic similarity of derivatives and bases, but the meaning of the morphological category in question itself also plays a role.

References:

- Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press.
- Barker, Chris. 1998. Episodic *-ee* in English: A Thematic Role Constraint on New Word Formation. *Language* 74(4). 695.
- Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford Univ. Press.
- Bojanowski, Piotr, Edouard Grave, Armand Joulin & Tomas Mikolov. 2016. Enriching Word Vectors with Subword Information. arXiv preprint arXiv:1607.04606.
- Haspelmath, Martin. 2001. Word Classes and Parts of Speech. In Neil J. Smelser & Paul B. Baltes (eds.), *International encyclopedia of the social & behavioral sciences*, 16538–16545. Amsterdam: Elsevier.
- Hay, Jennifer & Harald Baayen. 2003. Phonotactics, parsing and productivity. *Italian Journal of Linguistics* 1. 99–130.
- Huyghe, Richard & Marine Wauquier. 2020. What's in an agent? *Morphology* 30(3). 185–218.
- Kawaletz, Lea. 2021. *The semantics of English -ment nominalizations*. PhD Dissertation, Heinrich-Heine-Universität Düsseldorf.
- Lapesa, Gabriella, Lea Kawaletz, Ingo Plag, Marios Andreou, Max Kisselew & Sebastian Padó. 2018. Disambiguation of newly derived nominalizations in context: A Distributional Semantics approach. *Word Structure* 11(3). 277–312.
- Mikolov, Tomas, Edouard Grave, Piotr Bojanowski, Christian Puhersch & Armand Joulin. 2018. *Advances in Pre-Training Distributed Word Representations*.

Plag, Ingo. 2004. Syntactic category Information and the semantics of derivational morphological rules. *Folia Linguistica* 38(34). 193–225.

Plag, Ingo. 2018. *Word-formation in English* (Cambridge textbooks in linguistics). Cambridge, United Kingdom, New York, NY: Cambridge University Press.

Plag, Ingo, Marios Andreou & Lea Kawaletz. 2018. A frame-semantic approach to polysemy in affixation. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology*, 467–486. Berlin: Language Science Press.

Schneider, Viktoria. 2022. Eventualities in the semantics of denominal nominalizations. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*. de Gruyter. Accepted.

Szabó, Zoltán G. 2015. Major Parts of Speech. *Erkenntnis* 80(S1). 3–29. Tomaschek, Fabian, Peter Hendrix & R. H. Baayen. 2018. Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics* 71. 249–267.

Van Valin, Robert D. & Randy J. LaPolla. 2002. *Syntax: Structure, meaning and function* (Cambridge textbooks in linguistics). Cambridge: Cambridge Univ. Press.

Wauquier, Marine, Cécile Fabre & et N. Hathout. 2018. Différenciation sémantique de dérivés morphologiques à l'aide de critères distributionnels. *SHS Web of Conferences* 46. 8006.