

Measuring Semantic Content To Assess Asymmetry in Derivation

Sebastian Padó* Alexis Palmer* Max Kisselew* Jan Šnajder†

*Institut für maschinelle Sprachverarbeitung, Stuttgart University, Germany

†Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

{kisselmx,pado,palmeras}@ims.uni-stuttgart.de jan.snajder@fer.hr

Motivation and theoretical perspective. Morphological derivation is the process of forming new words (**derived terms**) from existing ones (**base terms**) such as nominalization. E.g., the English noun *dancer* is derived from the verb *dance* by applying the agentive suffix *-er*. In contrast to inflection, which changes the grammatical properties of the base term, derivation arguably creates a new lexical item. Derivation combines (typically regular) surface changes (add *-er*) with semi-regular semantic shifts (*-er* can refer to, among others, agents and instruments of the respective verbs), and has received interest in theoretical linguistics (Plank, 1981) and recently in computational linguistics (Lazaridou et al., 2013).

In this study, our goal is to investigate a theoretical claim about derivation with CL methods. The claim is that derivation is an inherently **directional** process with respect to meaning. Laca (2001) claims that “a derived lexeme presupposes the lexeme it is derived from,” suggesting that the meanings of derived terms involve additional meaning components. Being a *dancer* presupposes a *dancing* event, as well as additional, more specific information about the relation to the event. This forms our hypothesis:

(1) **Hypothesis:** Derived terms have **more semantic content** than their respective base terms.

Metrics from information theory. We investigate this hypothesis with corpus-based distributional models. The main methodological question is how to operationalize the notion of “semantic content”. We experiment with two information-theoretic measures proposed by recent studies.

1. Santus et al. (2014) proposed **entropy** to determine the semantic generality of words based on their distributional vectors. We take the entropy of the context word distribution for a term directly as a measure of the information content of the term itself. Our hypothesis is that semantic content is *inversely* correlated with entropy, since higher semantic content corresponds to more restrictions on context, which means lower entropy.
2. Herbelot and Ganesalingam (2013) computed the **Kullback-Leibler (KL) divergence** between distributional vectors and the prior distribution of context words (the “neutral” context vector, i.e. the centroid of all words) as a measure of the vector’s semantic content. This intuition carries over directly to our scenario: we assume higher KL divergence corresponds to higher semantic content.

Our testbed: German derivational morphology. The data for this study comes from DERivBase (Zeller et al., 2013), a large-coverage resource for German derivational morphology. DERivBase covers 280k lemmas and was created from a rule-based framework. Each base-derived lemma pair in DERivBase is labeled with a sequence of derivation patterns from a set of 267 patterns, allowing for easy extraction of word pairs for individual derivational patterns. We consider six within-POS derivation patterns in German (see Table 1), extracting 80 word pairs for each pattern and requiring that both base and derived lemmas appear at least 20 times in SdeWaC, a 800M-word German web corpus (Faaß and Eckart, 2013).

We build a vector space from SdeWaC, part-of-speech tagged and lemmatized using TreeTagger (Schmid, 1994) with back off for unrecognized words to lemmatization from the MATE tools (Bohnet, 2010) and handling of separated prefix verbs based on MATE dependency analysis. We use the 10K most frequent content words (nouns, verbs, adjectives) in SdeWAC as contexts and use simple, L1-normalized count vectors. The neutral vector is computed as the centroid of all count vectors.

Given this set of derivational patterns, and the distributional vectors associated with them, we now ask whether we find consistent effects.

ID	Pattern	Base	Derived	Sample word pair	English translation	Entropy	KL
AA02	<i>un-</i>	adj	adj	<i>sagbar</i> → <i>unsagbar</i>	<i>sayable</i> → <i>unspeakable</i>	60/20	62/18
AA03	<i>anti-</i>	adj	adj	<i>religiös</i> → <i>antireligiös</i>	<i>religious</i> → <i>antireligious</i>	78/2	78/2
NN02	<i>-in</i>	noun	noun	<i>Bäcker</i> → <i>Bäckerin</i>	<i>baker</i> → <i>female baker</i>	76/4	74/6
NN57	<i>-chen</i>	noun	noun	<i>Schiff</i> → <i>Schiffchen</i>	<i>ship</i> → <i>small ship</i>	74/6	75/5
VV13	<i>an-</i>	verb	verb	<i>backen</i> → <i>anbacken</i>	<i>to bake</i> → <i>to stick, burn</i>	71/9	68/12
VV31	<i>durch-</i>	verb	verb	<i>atmen</i> → <i>durchatmen</i>	<i>to breathe</i> → <i>to breathe deeply</i>	76/4	75/5

Table 1: Derivation patterns with base and derived categories, representative examples (and translations). Right-hand side columns: number of word pairs which match/mismatch the hypothesis.

Results for entropy. The distributions of entropies for each derivational pattern are shown on the left in Figure 1. Even though pattern averages differ substantially, and there is much lexical variation, our hypothesis (lower entropy for derived terms) holds for about 90% of the 480 word pairs. Interestingly, there is little variation across parts of speech. Two patterns that stick out due to substantially lower rates are the adjectival negation suffix *un-* and the verbal prefix *an-*. For *un-*, the counterexamples fall into two broad groups. One group consists of cases where the derived term seems to be the more basic term (*unentbehrlich* (*indisposable*) – *entbehrlich* (*disposable*)). In the other, derived terms undergo a meaning shift beyond negation (*unumwunden* (*honestly, straightforwardly*) – *umwunden* (*wound up*)). The negative cases for *an-* are dominated by the second category, e.g., *kündigen* (*cancel*) – *ankündigen* (*announce*)).

Results for KL divergence. The distributions of KL divergences are shown on the right hand side of Figure 1. Although numerically different and to be interpreted inversely (derived terms should have higher divergence), the results are very similar to entropy – not surprising, given their close relationship. The similarity extends from the distributions to the percentage of lemma pairs supporting our hypothesis, which is again around 90%. Notably, the same two patterns fall out as problematic, and there is very high overlap between the two. Interestingly, KL divergence tends to pick out cases of the second category (additional meaning shifts) rather than the first category (derivational terms that are basic).

Outlook. In this abstract, we report on a distributional study to assess the impact of morphological derivation on semantic content. Using two information-theoretic measures to quantify semantic content, we find very strong empirical evidence for asymmetry: derived terms indeed have more semantic content. While stable across parts of speech and, notably, across word pairs, we also find hints of the notorious semi-regularity in morphology. Two of our six patterns frequently give rise to derivations that undergo additional semantic shifts. Our next step will be to analyse why specifically these patterns and word pairs behave unpredictably, considering theories that relate meaning shift and frequency (Haspelmath, 2008).

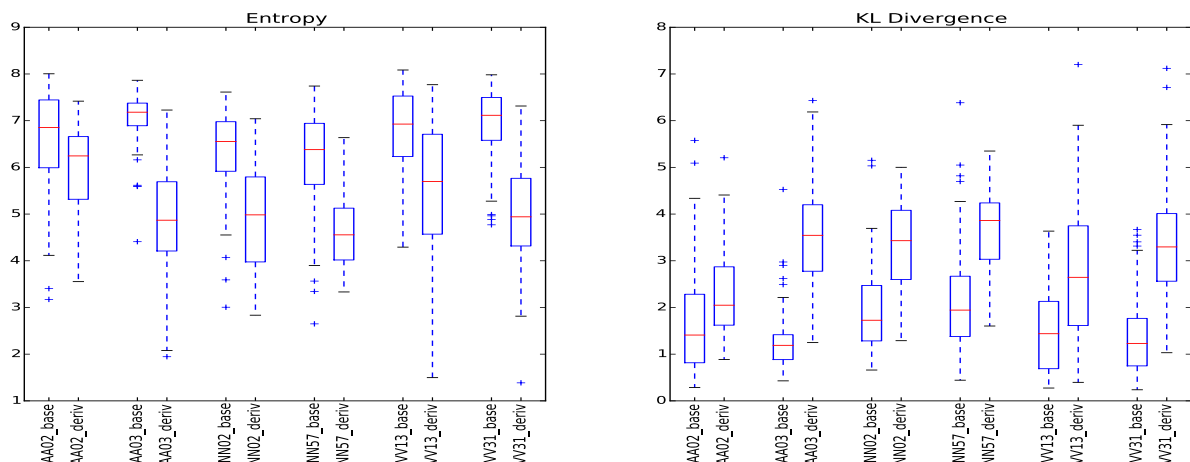


Figure 1: (a) Entropy and (b) KL divergence (compared to neutral distribution) of base and derived terms

References

- Bohnet, B. (2010). Top accuracy and fast dependency parsing is not a contradiction. In *Proceedings of COLING*, Beijing, China.
- Faaß, G. and K. Eckart (2013). SdeWaC – a corpus of parsable sentences from the web. In *Language Processing and Knowledge in the Web*, Lecture Notes in Computer Science. Springer.
- Haspelmath, M. (2008). Creating economical morphosyntactic patterns in language change. In J. Good (Ed.), *Language universals and language change*, pp. 185–214. OUP.
- Herbelot, A. and M. Ganesalingam (2013). Measuring semantic content in distributional vectors. In *Proceedings of ACL*, pp. 440–445.
- Laca, B. (2001). Derivation. In M. Haspelmath, E. König, W. Oesterreicher, and W. Raible (Eds.), *Language Typology and Language Universals: An International Handbook*, Volume 1, pp. 1214–1227. Berlin: Walter de Gruyter.
- Lazaridou, A., M. Marelli, R. Zamparelli, and M. Baroni (2013). Compositional-ly derived representations of morphologically complex words in distributional semantics. In *Proceedings of ACL*, Sofia, Bulgaria.
- Plank, F. (1981). *Morphologische (Ir-)Regularitäten. Aspekte der Wortstrukturtheorie*. Tübingen: Narr.
- Santus, E., A. Lenci, Q. Lu, and S. S. Im Walde (2014). Chasing hypernyms in vector spaces with entropy. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, Volume 2, pp. 38–42.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of NeMLaP*, Manchester, UK.
- Zeller, B., J. Šnajder, and S. Padó (2013). DERivBase: Inducing and evaluating a derivational morphology resource for German. In *Proceedings of ACL*, Sofia, Bulgaria.