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

Funded by Deutsche Forschungsgemeinschaft (DFG SFB 732, B9)

1. Morphological Derivation

- The process of forming new words (derived terms) from existing ones (base terms) dance+er ⇒ dancer
- Combines surface changes with semi-regular semantic shifts
- Theoretical claim: inherently **directional** process with respect to meaning (Laca, 2001) dancer presupposes dancing event, relational information
  - Our hypothesis: derived terms have **more semantic content** than their respective base terms
  - Our goal: measure semantic content from corpus data and assess hypothesis

2. Measuring Semantic Content

Operationalized in distributional semantic framework, using two metrics from information theory

Entropy (H)

- Santus et al. (2014): entropy of distributional vectors as measure of semantic generality of words
- Here: entropy of a term's vector as measure of information content
- Entropy computed for both base and derived terms
- High semantic content ⇒ low entropy

KL Divergence (D)

- Herbelot and Ganesalingam (2013): KL divergence between term vector and "neutral" context vector as measure of semantic content
- Here: "neutral" vector computed as centroid vector for all words
- Both base and derived vectors compared to centroid vector
- High semantic content ⇒ high KL divergence from neutral vector

Two metrics not equivalent; D incorporates both cross-entropies and entropy difference: $D(d||n) - D(b||n) = \big(H(d,n) - H(b,n)\big) - \big(H(d) - H(b)\big)$

3. Data

- Lemmatized, POS-tagged SdeWaC (Faaß & Eckart, 2013)
- 10K most frequent content words as contexts
- Count vectors, L1-normalized

Derivational patterns and word pairs

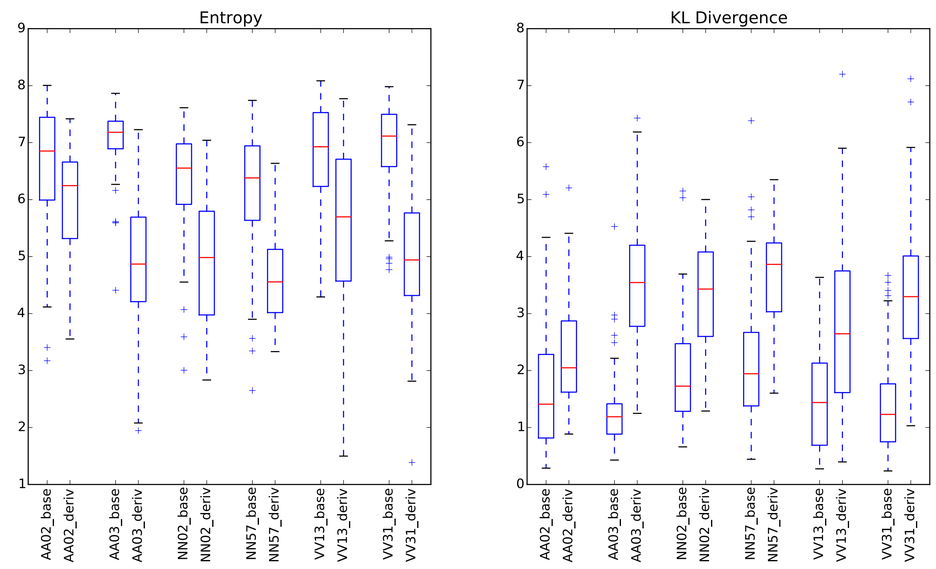
- From DErivBase (Zeller et al., 2013)
- Two each of A-A, N-N, V-V patterns
- 80 word pairs per pattern, corpus frequency ≥ 20

ID	Pattern	Sample word pair	English translation
AA02	*un-*	sagbar → unsagbar	sayable → unspeakable
AA03	*anti-*	religiös → antireligiös	religious → antireligious
NN02	*-in*	Bäcker → Bäckerin	baker → female baker
NN57	*-chen*	Schiff → Schiffchen	ship → small ship
VV13	*an-*	backen → anbacken	to bake → to stick, burn
VV31	*durch-*	atmen → durchatmen	to breathe → to breathe deeply

4. Results

Expectations and outcomes

1. **Entropy:** entropy of base terms is higher than that of derived terms
2. **KL divergence:** base terms show lower KL divergence (compared to the neutral vector) than do derived terms



Assessing the hypothesis

- Results strongly support hypothesis, across parts of speech
- Roughly 90% of word pairs conform to expectations

Metric	A: *un-*	A: *anti-*	N: *-in*	N: *-chen*	V: *an-*	V: *durch-*
Entropy	60/20	78/2	76/4	74/6	71/9	76/4
KL	62/18	78/2	74/6	75/5	68/12	75/5

Table: For each pattern, number of word pairs which match/mismatch the hypothesis

- Two main types of counterexamples:
  1. derived term is more basic
     entbehrlich (disposable) ⇒ *unentbehrlich (indisposable)*
  2. derived term undergoes additional meaning shift
     kündigen (cancel) ⇒ *ankündigen (announce)*
- Entropy finds more cases of first type; KL, more of second type
- Mixed-effects logistic regression analysis shows
  - highly significant effect of derivational status (+derived ⇒ +semantic_content)
  - additional substantial effects of frequency (+freq ⇒ −semantic_content)
  - no effect of POS

5. Conclusion

- Very strong empirical evidence for asymmetry: derived terms indeed have more semantic content than base terms
- Non-conforming word pairs show evidence of morphological semi-regularity (additional semantic shifts)
- Next: further investigate misbehaving patterns and word pairs, considering e.g. relationship between meaning shifts and frequency (Haspelmath, 2008)

Also see our poster tomorrow at IWCS!

References

- Faaß, G. and K. Eckart (2013). SdeWaC – A corpus of parsable sentences from the web. Language Processing and Knowledge in the Web.
- Haspelmath, M. (2008). Creating economical morphosyntactic patterns in language change. In *Language universals and language change*. OUP.
- Herbelot, A. and M. Ganesalingam (2013). Measuring semantic content in distributional vectors. Proceedings of ACL.
- Kisselew, M., S. Padó, A. Palmer, and J. Šnajder (2015). Obtaining a Better Understanding of Distributional Models of German Derivational Morphology. Proceedings of IWCS.
- Laca, B. (2001). Derivation. In *Language Typology and Language Universals: An International Handbook.* Volume 1. Walter de Gruyter.
- Santus, E., A. Lenci, Q. Lu, and S. S. Im Walde (2014). Chasing hypernyms in vector spaces with entropy. Proceedings of EACL.
- Zeller, B., J. Šnajder, S. Padó (2013). DErivBase: Inducing and evaluating a derivational morphology resource for German. Proceedings of ACL.