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

1. Morphological Derivation

- The process of forming new words (derived terms) from existing ones (base terms)
- Combines surface changes with semi-regular semantic shifts

Theoretical claim: inherently directional process with respect to meaning (Laca, 2001)
- 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: \( D(d||n) - D(b||n) = H(d,n) - H(b,n) - (H(d) - H(b)) \)

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

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>60/20</td>
<td>78/2</td>
<td>76/4</td>
<td>74/6</td>
<td>71/9</td>
<td>76/4</td>
</tr>
<tr>
<td>KL</td>
<td>62/18</td>
<td>78/2</td>
<td>74/6</td>
<td>75/5</td>
<td>68/12</td>
<td>75/5</td>
</tr>
</tbody>
</table>

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

- Two main types of counterexamples:
  1. derived term is more basic
  2. derived term undergoes additional meaning shift

- Mixed-effects logistic regression analysis shows
  - highly significant effect of derivational status
  - additional substantial effects of frequency
  - 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)

References