Predicting the Direction of Derivation in English Conversion

Max Kisselew\(^*\), Laura Rimell\(^\dagger\), Alexis Palmer\(^\ddagger\), Sebastian Padó\(^*\)

\(^*\) IMS, Stuttgart University, Germany \(^\dagger\) Computer Laboratory, University of Cambridge, UK

\(|\) Leibniz ScienceCampus, IOL, Heidelberg University, Germany

1. Morphology and Direction of Conversion

- **Conversion** changes grammatical category of a word without overt morphological marking, e. g.: `tunnel (n.) \(\rightarrow\) tunnel (v.), walk (v.) \(\rightarrow\) walk (n.)`
- Various theoretical accounts of conversion: Uncategorized roots (underspecification) vs. directed derivation

<table>
<thead>
<tr>
<th>Lexicon</th>
<th>Surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>fish (n.)</td>
<td>fish (n.)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Underspecification</th>
<th>Directionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>fish (n.)</td>
<td>fish (v.)</td>
</tr>
</tbody>
</table>

Research Question

In a corpus-based study, which factors are able to account for diachronic precedence in cases of English V-to-N and N-to-V conversion?

2. Hypotheses

1. Derived forms are **less frequent** than their bases (Harwood and Wright, 1956; Hay, 2001)
2. Derived forms are **more semantically specific** than their bases (Koontz-Garboden, 2007; Plag, 2003), as approximated by information theoretic measures

3. Data

- **Gold standard**: Historical precedence data from CELEX (Baayen et al., 1995) for English
  - 1,044 monomorphic English N-to-V lemma pairs
  - 948 monomorphic English V-to-N lemma pairs
- **Corpus**: Concatenation of the lemmatized and part-of-speech (PoS) tagged BNC and ukWaC corpora containing 2.36 billion tokens
- **Semantic vector space**: Separate vectors c.noun and c.verb for each conversion case c
  - BOW count vectors, 10000 dimensions, context window \( \pm 5 \)
  - Downsampling: For each verb-noun conversion pair, both vectors are constructed from the same number of occurrences

4. Specificity Measures

- Two measures for semantic specificity of a word:
  - Entropy:
    \[
    H(v) = -\sum_{i} v_i \cdot \log(v_i) 
    \]
    (high semantic specificity \(\sim\) low entropy)
  - Kullback-Leibler (KL) divergence:
    \[
    D(v||n) = \sum_{i} v_i \cdot \log \left( \frac{v_i}{n_i} \right) 
    \]
    (high semantic specificity \(\sim\) high KL divergence from neutral vector)
  - KL divergence between term vector and “neutral” context vector \(n\) as a measure of the vector’s semantic specificity
  - Here: “neutral” vector \(n\) computed as centroid vector for all words in the corpus

5. Experiments

- **Testing hypothesis 1 (Frequency)**:
  - If \( f(N) > f(V) \) then N-to-V (else V-to-N)
- **Testing hypothesis 2 (Semantic specificity)**:
  - If \( H(N) > H(V) \) then N-to-V (else V-to-N)
  - If \( D(V||n) > D(N||n) \) then N-to-V (else V-to-N) (where \(n\) is the neutral vector)
- Combined model: combination of individual indicators (standardized differences in log frequency, entropy, and KL divergence within each pair) as features in a logistic regression model

6. Results

<table>
<thead>
<tr>
<th>Predictor</th>
<th>N-to-V</th>
<th>V-to-N</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most Frequent Class</td>
<td>100%</td>
<td>0%</td>
<td>52.4%</td>
</tr>
<tr>
<td>Entropy (H)</td>
<td>50.1%</td>
<td>75.5%</td>
<td>62.2%</td>
</tr>
<tr>
<td>KL divergence</td>
<td>53.8%</td>
<td>76.7%</td>
<td>64.6%</td>
</tr>
<tr>
<td>Frequency</td>
<td>84.7%</td>
<td>58.7%</td>
<td>72.3%</td>
</tr>
<tr>
<td>Freq + (H + KL)</td>
<td>77.4%</td>
<td>76.0%</td>
<td>76.8%</td>
</tr>
</tbody>
</table>

Accuracies for predicting the direction of derivation

- Large difference in results between N-to-V and V-to-N
- Frequency best predictor for N-to-V cases
  - Large variety in meaning shifts
  - Verb describes an ‘action having to do with the noun’. E. g.: `celluloid the door open`, meaning ‘use a credit card to spring the lock open’ (Clark and Clark, 1979)
  - Regular semantics of conversion
  - Specificity predictors better for V-to-N cases
  - Noun is likely to refer to the event described by the verb or its result (Grimshaw, 1990)
  - More regular semantics of conversion
  - Simple combination does well for both cases

7. Discussion and Conclusion

- **Striking complementarity in the ability of frequency and semantic specificity to account for the direction of conversion in N-to-V and V-to-N cases**
- **Frequency best predictor for N-to-V cases**
  - Large variety in meaning shifts
  - Verb describes an ‘action having to do with the noun’. E. g.: `celluloid the door open`, meaning ‘use a credit card to spring the lock open’ (Clark and Clark, 1979)
  - Regular semantics of conversion
  - Specificity predictors better for V-to-N cases
  - Noun is likely to refer to the event described by the verb or its result (Grimshaw, 1990)
  - More regular semantics of conversion
  - Simple combination does well for both cases

Acknowledgements

MK and SP acknowledge partial funding by Deutsche Forschungsgemeinschaft (SFB 732, Project B9), LR acknowledges EPSRC grant EP/I037512/1 and ERC Starting Grant DisCoTex (306920). AP acknowledges Leibniz Association grant SAS-2015-IDS-LWC and the Ministry of Science, Research, and Art of Baden-Württemberg. MK and SP acknowledge partial funding by Deutsche Forschungsgemeinschaft (SFB 732, Project B9), LR acknowledges EPSRC grant EP/I037512/1 and ERC Starting Grant DisCoTex (306920). AP acknowledges Leibniz Association grant SAS-2015-IDS-LWC and the Ministry of Science, Research, and Art of Baden-Württemberg.

References


