Problems of DURel annotation measures for semantic change

March 9, 2021

Dominik Schlechtweg
Supervisor: apl. Prof. Dr. Sabine Schulte im Walde
Institute for Natural Language Processing, University of Stuttgart, Germany
Annotation of Lexical Semantic Change

- **Diachronic Usage Relatedness (DURel)** (Schlechtweg, Schulte im Walde, & Eckmann, 2018)
- five annotators
- annotate sentence pairs from German diachronic DTA corpus for degree of semantic relatedness
- we measure
  - **innovative meaning change**: emergence of a full-fledged additional meaning of a word; old and new meaning are related by polysemy
  - **reductive meaning change**: loss of a full-fledged meaning of a word

(Blank, 1997)
4: Identical
3: Closely Related
2: Distantly Related
1: Unrelated
0: Cannot decide

Table 1: Four-point scale of relatedness derived from Brown (2008).
Figure 1: 2-dimensional use spaces Zlatev (2003) in two time periods with a target word $w$ undergoing innovative meaning change. Dots represent uses of $w$. Spatial proximity of two uses means high relatedness.
Two Measures of Lexical Semantic Change

i) $\Delta_{\text{LATER}}(w) = \text{Mean}_{\text{later}}(w) - \text{Mean}_{\text{earlier}}(w)$

- measures changes in the degree of mean relatedness of words
- positive vs. negative values on this measure indicate innovative vs. reductive meaning change.
- is justified by the observation that lexical semantic change is strongly correlated with polysemy (Blank, 1997)
- collapses where innovation and reduction occur together
Two Measures of Lexical Semantic Change

Figure 2: $\Delta$LATERT: Rank of target words.
Problem: Multiple Changes

Figure 3: Innovative followed by reductive meaning change.
Two Measures of Lexical Semantic Change

ii) $\text{COMPARE}(w) = \text{Mean}_{\text{compare}}(w)$

- directly measures the relatedness between EARLIER and LATER
- High vs. low values on this measure indicate weak vs. strong change
- is justified by the idea that emerging meanings show up as uses which are different from the old meaning
- collapses where words are polysemous (confuses polysemy and change)
Figure 4: Judgment frequencies of *Presse*. $\Delta_{\text{LATER}}$ wrongly predicts no change, COMPARE strong change.
Two Measures of Lexical Semantic Change

Figure 5: Judgment frequencies of **Feder**. $\Delta_{\text{LATER}}$ correctly predicts no change, **COMPARE** strong change.
Normalization of `COMPARE`

1. $\Delta \text{COMPARE}(w) = \text{Mean}_{\text{compare}}(w) - \text{Mean}_{\text{earlier}}(w)$
   - measures how much the relatedness between EARLIER and LATER **exceeds the relatedness in** EARLIER
   - high values on this measure mean strong reduction
   - low values mean meaning innovation or difference in use
   - innovation and reduction will show up as negative versus positive values
   - reduction will only be predicted if an old meaning is preserved
   - has sampling problems
Figure 6: $\Delta$COMPARE: Rank of target words.
Normalization of COMPARE

Figure 7: Judgment frequencies of **Vorwort**. ΔLATER wrongly predicts no change, COMPARE strong change. *(See also case of **Presse**).*
Problem: Different Sampling Strategies for \textit{EARLIER} and \textit{COMPARE}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{minimal_sampling_example.png}
\caption{Minimal sampling example.}
\end{figure}
Bibliography


