Diachronic Usage Relatedness (DURel): A Framework for the Annotation of Lexical Semantic Change

Dominik Schlechtweg, Sabine Schulte im Walde, Stefanie Eckmann

University of Stuttgart
Germany

Introduction

- major obstacle in the computational modeling of semantic change is evaluation
- no reliable test set of semantic change for any language
- we counteract this lack of resources by extending a framework of synchronic polysemy annotation to the annotation of **Diachronic Usage Relatedness (DURel)**
- creating the first test set of lexical semantic change for German

Related Work

- Blank (1997) develops criteria to distinguish the relatedness of use pairs in the context of lexical semantic change
- various graded polysemy annotation studies of use pairs on relatedness (or similar) scales (Brown, 2008; Erk, McCarthy, & Gaylord, 2013; Soares da Silva, 1992)

Annotation Scale

4: Identical
3: Closely Related
2: Distantly Related
1: Unrelated

0: Cannot decide

Table 1: Our 4-point scale of relatedness derived from Brown (2008).

References

Blank, A. (1997). *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen.* Tübingen: Niemeyer.

Brown, S. W. (2008). Choosing sense distinctions for WSD: Psycholinguistic evidence. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers* (pp. 249–252). Stroudsburg, PA, USA.

Erk, K., McCarthy, D., & Gaylord, N. (2013). Measuring word meaning in context. *Computational Linguistics*, *39*(3), 511–554.

Soares da Silva, A. (1992). Homonímia e polissemia: Análise sémica e teoria do campoléxico. In *Actas do xix congreso internacional de lingüística e filoloxía románicas* (Vol. 2, pp. 257–287). La Coruña: Fundación Pedro Barrié de la Maza.

Lexical Semantic Change

Blank (1997) distinguishes two main types of lexical semantic change:

- **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

Example of Innovative Meaning Change

EARLIER

(1) *An schrecklichen Donnerwettern und heftigen Regengüssen fehlt es hier auch nicht.*
'There is no lack of horrible thunderstorms and heavy rainstorms.'

LATER

(2) a. *Oder es überschauerte ihn wie ein Donnerwetter mit Platzregen.*
'Or he was doused like a thunderstorm with a heavy shower.'
b. *Potz Donnerwetter!*
'Man alive!"

Diachronic Semantic Relatedness

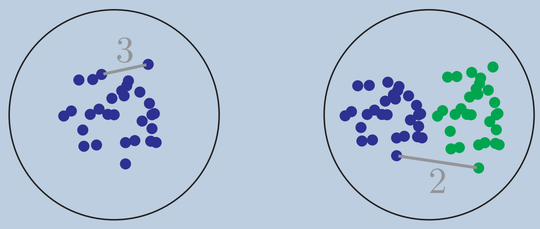

Figure 4: 2-dimensional use spaces in two time periods with a target word w undergoing innovative meaning change. Dots represent uses of w. Spatial proximity of two uses means high relatedness.

- **basic idea**: we measure the mean semantic relatedness of use pairs of a word w over time
  $\Delta\text{LATER}(w) = Mean_{ltr.}(w) - Mean_{erl.}(w)$
  increase vs. **decrease** indicate reductive vs. innovative meaning change
- to capture complex constellations we compare uses from EARLIER and LATER directly:
  $\text{COMPARE}(w) = Mean_{cmp.}(w)$
  high vs. **low** values indicate weak vs. strong change

Annotation Study

- **five annotators** rated 1,320 German use pairs on relatedness scale in Table 1
- for **22 target words** we randomly sampled 20 use pairs per group from DTA corpus
- there are **three groups**: EARLIER (1750-1800), LATER (1850-1900) and COMPARE
- order within pairs was randomized, pairs from all groups were mixed and randomly ordered

	1	2	3	4	5
1		0.59	0.63	0.67	0.66
2			0.57	0.64	0.65
3				0.64	0.62
4					0.68
avg	0.71	0.68	0.68	0.75	0.74

Table 2: Correlation matrix for pairwise correlation agreement of annotators

Results

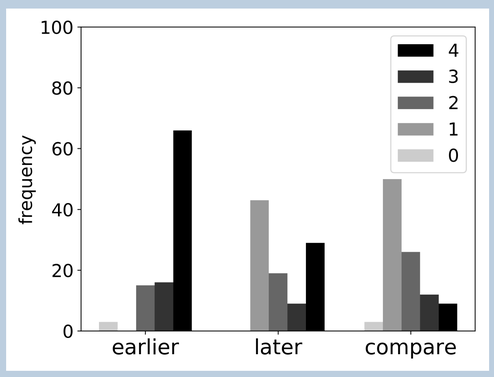

Figure 1: Judg. freq. for *Donnerwetter* (innovative).

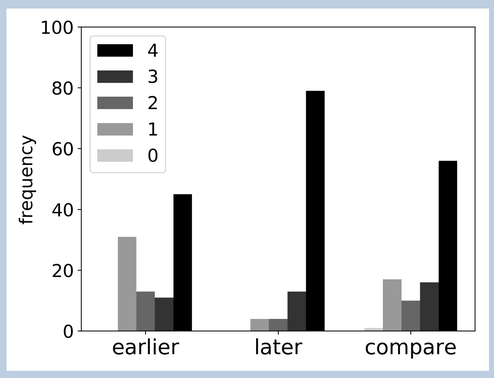

Figure 2: Judgment freq. for *Zufall* (reductive).

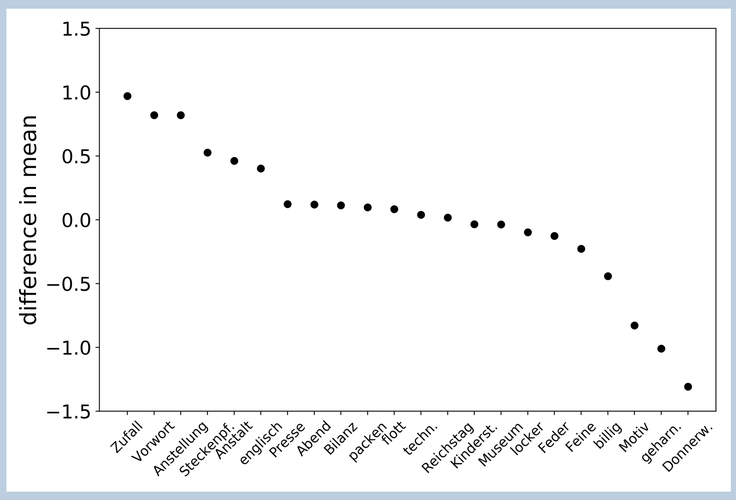

Figure 3: ΔLATER: Rank of target words.

ΔLATER:

- three lowermost words are innovative, three topmost words are reductive meaning changes
- mean value for reduction is 0.39, while it is -0.18 for innovation
- overall distinguishes well between innovation and reduction
- should be used only for **simple constellations**

COMPARE:

- does not distinguish between innovation and reduction (low values can be both)
- should be used only for **monosemous words**

Discussion

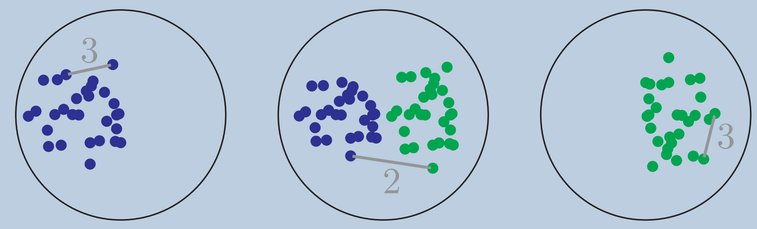

Figure 5: Innovative followed by reductive meaning change. ΔLATER predicts no change, while COMPARE predicts change.

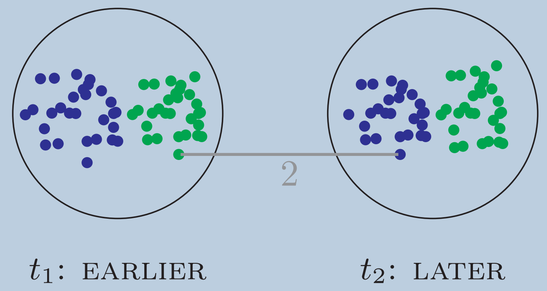

Figure 6: Polysemous semantically stable word. ΔLATER predicts no change, while COMPARE predicts change.

Preliminary solution:
$\Delta\text{COMPARE}(w) = Mean_{cmp.}(w) - Mean_{erl.}(w)$