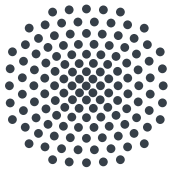


University of Stuttgart
Germany

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)

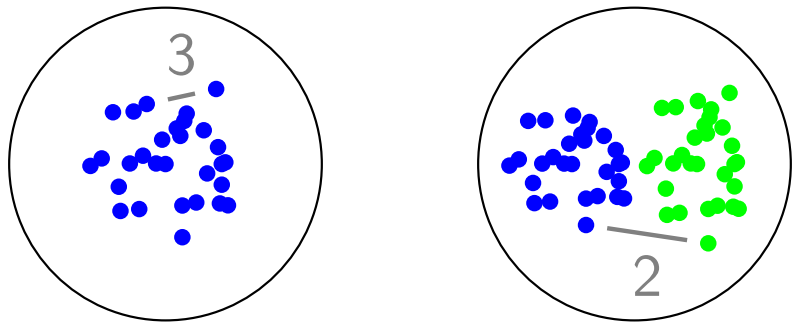
Scale

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

0: Cannot decide

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

Sampling



t_1: EARLIER t_2: LATER

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) $\Delta\textsc{later}(w) = Mean_{later}(w) - Mean_{earlier}(w)$

- 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

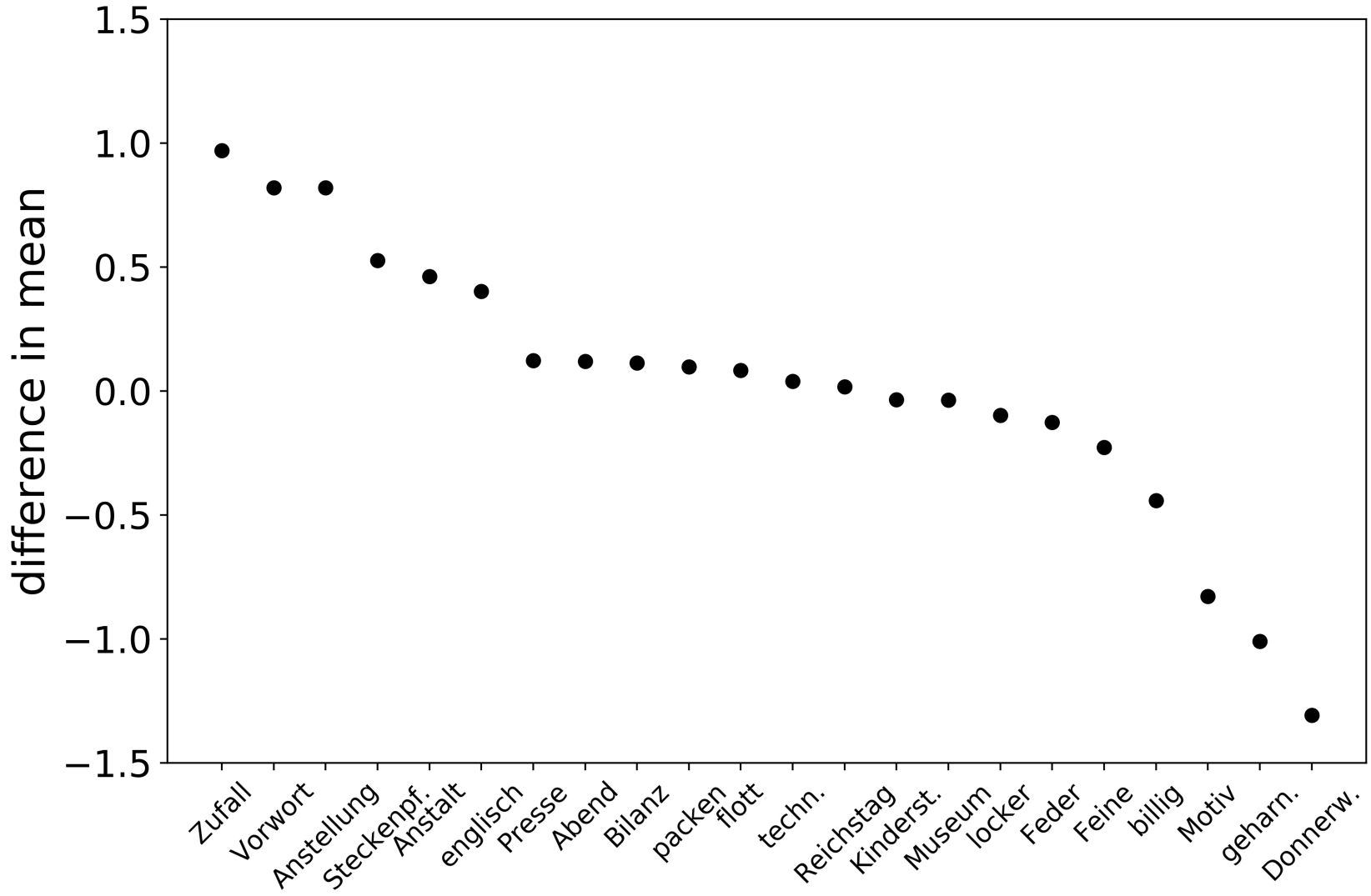
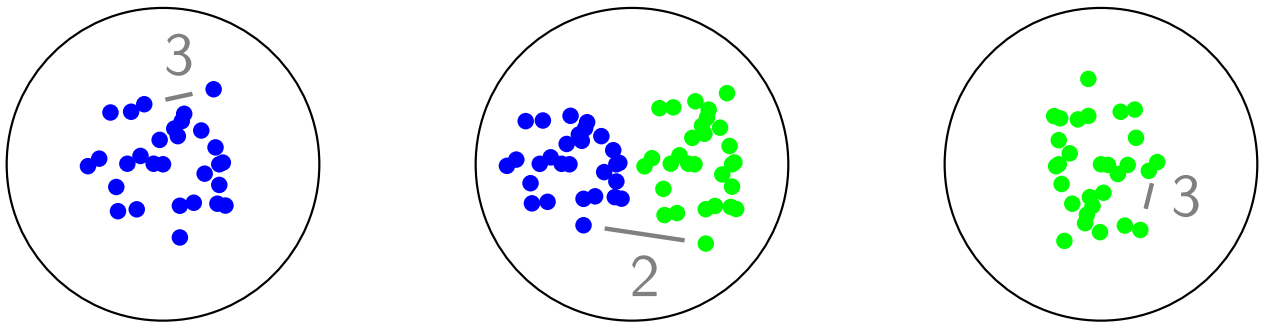


Figure 2: Δ_{LATER}: Rank of target words.

Problem: Multiple Changes



t_1: EARLIER t_2: LATER

Figure 3: Innovative followed by reductive meaning change.

Two Measures of Lexical Semantic Change

ii) $\textsc{compare}(w) = \mathit{Mean}_{\mathit{compare}}(w)$

- 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)

Two Measures of Lexical Semantic Change

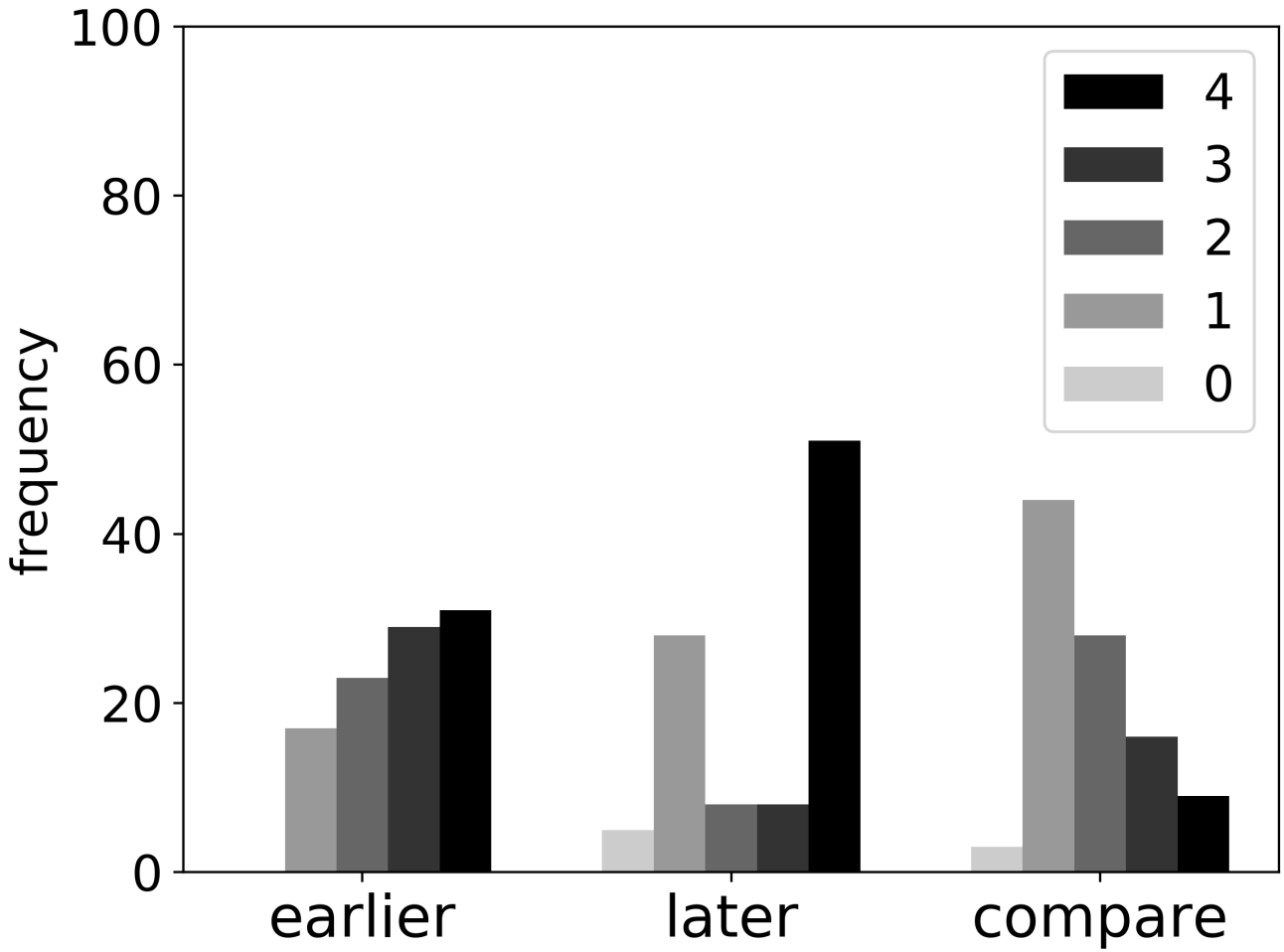


Figure 4: Judgment frequencies of **Presse**. ΔLATER wrongly predicts no change, COMPARE strong change.

Two Measures of Lexical Semantic Change

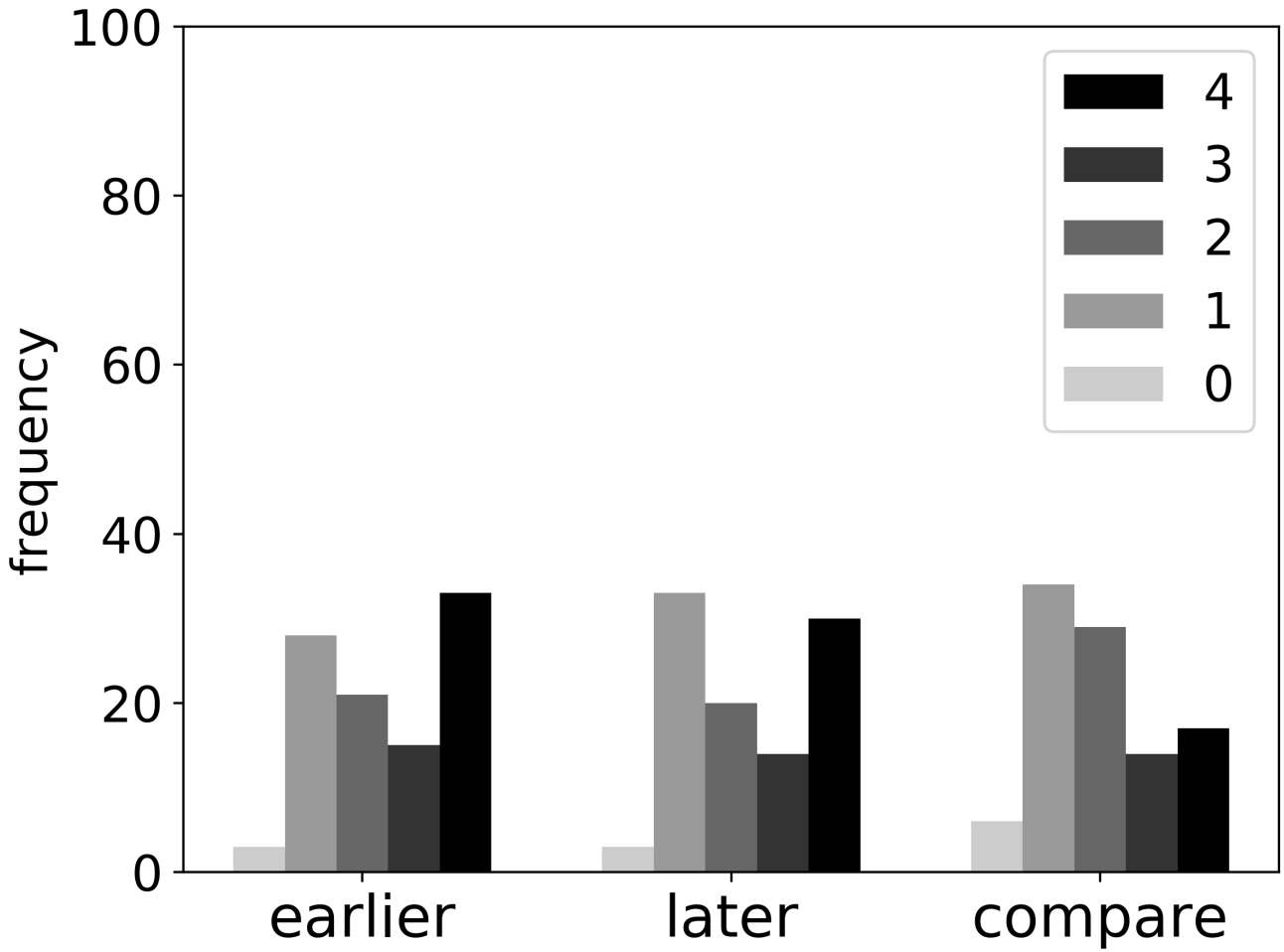


Figure 5: Judgment frequencies of **Feder**. Δ_{LATER} correctly predicts no change, COMPARE strong change.

Normalization of COMPARE

1. $\Delta\text{COMPARE}(w) = \mathit{Mean}_{\mathit{compare}}(w) - \mathit{Mean}_{\mathit{earlier}}(w)$
   - 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

Normalization of COMPARE

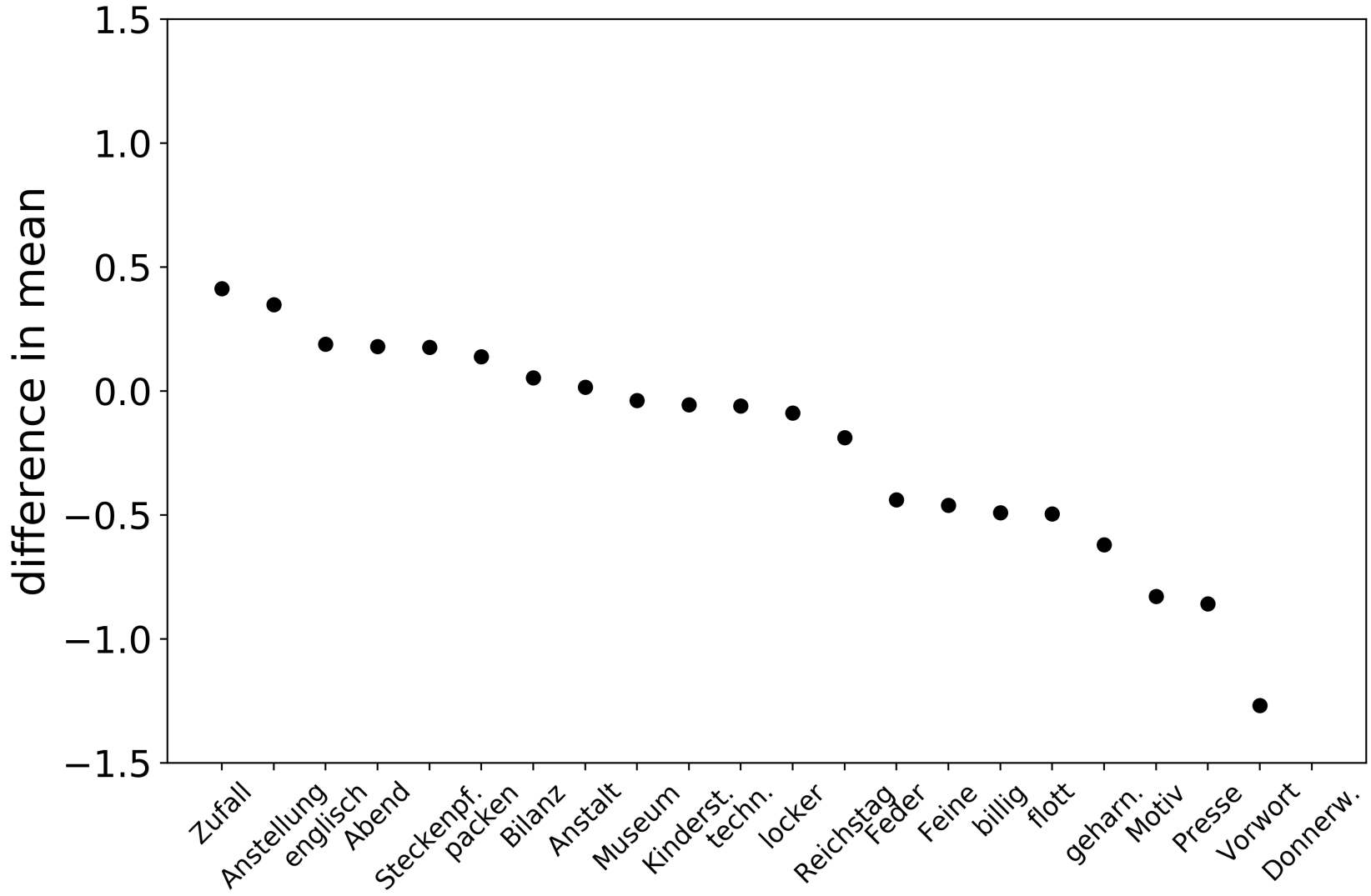


Figure 6: ΔCOMPARE: Rank of target words.

Normalization of COMPARE

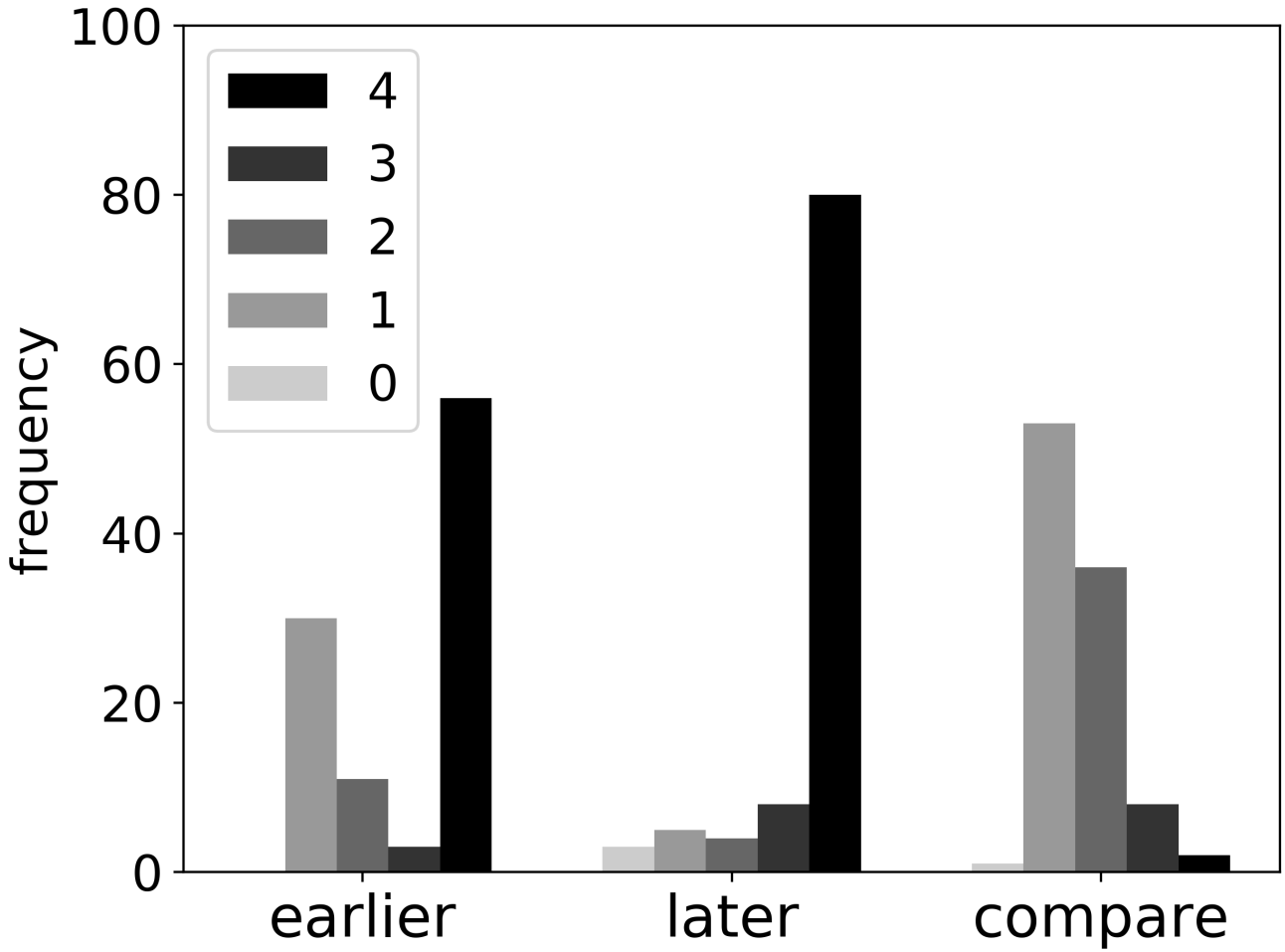


Figure 7: Judgment frequencies of **Vorwort**. ΔLATER wrongly predicts no change, COMPARE strong change. (See also case of **Presse**).

Problem: Different Sampling Strategies for EARLIER and COMPARE

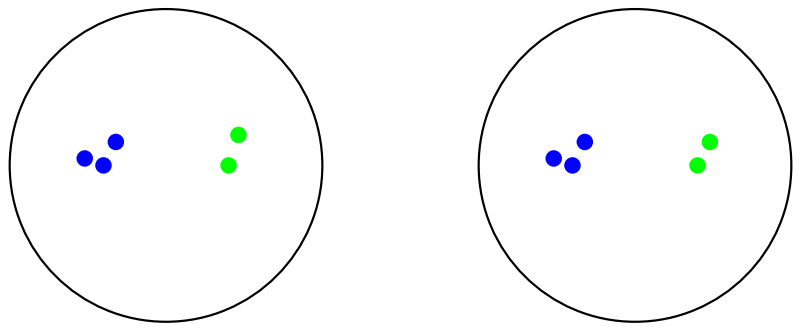


Figure 8: Minimal sampling example.

Bibliography

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.

Schlechtweg, D., Schulte im Walde, S., & Eckmann, S. (2018). Diachronic Usage Relatedness (DURel): A framework for the annotation of lexical semantic change. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 169–174). New Orleans, Louisiana.

Zlatev, J. (2003). Polysemy or generality? Mu. In H. Cuyckens, R. Dirven, & J. R. Taylor (Eds.), (Vol. 23, pp. 447–494). Mouton de Gruyter.