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).

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 Donnerwetter und heftigen Regensügen 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. Pots Donnerwetter!
‘Man alive!’

Diachronic Semantic Relatedness
• basic idea: we measure the mean semantic relatedness of use pairs of a word w over time $\Delta LATER(w) = Mean_{erl}(w) - Mean_{ cmp}(w)$
• 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

Results
Figure 1: Judg. freq. for Donnerwetter (innovative).
Figure 2: Judgment freq. for Zufall (reductive).
Figure 3: $\Delta$LATER: Rank of target words.

Discussion
Figure 5: Innovative followed by reductive meaning change. $\Delta$LATER predicts no change, while COMPARE predicts change.
Figure 6: Polysemous semantically stable word. $\Delta$LATER predicts no change, while COMPARE predicts change.

References

Table 2: Correlation matrix for pairwise correlation agreement of annotators