

---

# Comparing Annotation Frameworks for Lexical Semantic Change

---

Dominik Schlechtweg, Sabine Schulte im Walde  
*Institute for Natural Language Processing, University of Stuttgart*

dominik.schlechtweg@ims.uni-stuttgart.de, schulte@ims.uni-stuttgart.de

Evaluation in research on Lexical Semantic Change (LSC) detection is still an unsolved issue (cf. e.g. Lau et al., 2012; Cook et al., 2014; Frermann and Lapata, 2016; Takamura et al., 2017). This is mainly due to a lack of LSC resources, e.g. in the form of a test set comprising a number of words annotated for whether they changed semantically or not. In the creation of such a test set the same problems as in traditional word sense annotation are encountered (e.g. definition and dichotomy of word senses), as LSC is related to loss or emergence of word senses (cf. Blank, 1997, p. 113). In order to avoid these problems Schlechtweg et al. (2018) developed DUREl, an annotation framework relying on graded word sense annotation (Erk et al., 2013). However, as Schlechtweg et al. acknowledge, DUREl is unable to distinguish certain semantic constellations, e.g., it confuses polysemy with LSC. Although the authors propose a preliminary solution to this issue, DUREl can be shown to have a rather low correlation with true LSC.

In order to overcome the above-mentioned shortcomings of DUREl we propose an alternative annotation framework relying on the manual choice of representative uses for each word sense cluster. More specifically, we try to retrieve the underlying word sense distributions by choosing a representative centroid for each sense cluster against which a random sample of uses is to be compared. As we show by simulation of LSC in idealized graph models, from the inferred word sense distributions a more accurate value of LSC can be retrieved on the basis of a feasible sample size. At the same time the proposed framework retains the advantages of DUREl relying on graded word sense annotation of use pairs.

The two annotation strategies are compared and analyzed on the same set of words annotated by the same annotators. The newly proposed framework is not only relevant for research in LSC but for research on lexical semantics with relation to word sense distributions in general.

**References:** • Blank, A. (1997): *Pinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*. Tübingen: Niemeyer. • Cook, P., Lau, J. H., McCarthy, D. and Baldwin, T. (2014): Novel word-sense identification. In *Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers*, Dublin, Ireland, pp. 1624–1635. • Erk, K., McCarthy, D. and Gaylord, N. (2013): Measuring word meaning in context. *Computational Linguistics* 39(3), 511–554. • Frermann, L. and Lapata, M. (2016): A Bayesian model of diachronic meaning change. *Transactions of the Association for Computational Linguistics* 4, 31–45. • Lau, J. H., Cook, P., McCarthy, D., Newman, D. and Baldwin, T. (2012): Word sense induction for novel sense detection. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Stroudsburg, PA, USA, pp. 591–601. • Schlechtweg, D., Schulte im Walde, S. and 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*, New Orleans, Louisiana, pp. 169–174. • Takamura, H., Nagata, R. and Kawasaki, Y. (2017): Analyzing semantic change in Japanese loanwords. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, Valencia, Spain, pp. 1195–1204.