Empirical Studies of Word Sense Divergences across Language Varieties

Dominik Schlechtweg, Sabine Schulte im Walde

University of Stuttgart

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

Raum: ESA 1 Hauptgebäude (HG) Hörsaal (HS) K

Workshop description

Words change their senses not only over time but also across communities, domains, dialects, registers, and other language varieties (Wieling & Nerbonne 2015; Wiese & Pohle 2016; Del Tredici & Fernandez 2017; Ferrari et al. 2017; Hovy & Purschke 2018; Schlechtweg et al. 2019; i.a.).

An example for a diachronic sense divergence is the German noun *Vorwort*, which was mainly used in the meaning of „preposition" before ca. 1800 (Schlechtweg et al. 2018). Then it rapidly acquired a new meaning „preface", which after 1850 has nearly exclusively been used. An example for a synchronic domain-specific sense divergence is the German noun *Schnee* (Hätty et al. 2019). In general-language use, *Schnee* predominantly refers to „snow", while in the cooking domain the predominant meaning is the domain-specific „beaten egg whites". The German verb *heben* is an example for a **dialectal** lexical variation (Boberg et al. 2018), as it is used in the meaning „to lift" in standard German, while in the Southern-German dialect Swabian it is used in the meaning „to hold".

The above examples exhibit different predominant word senses with regard to specific language varieties. While each research field on language variety has its own tradition to explore word sense divergences, both from a theoretical and from an empirical perspective, this workshop aims to bring together interdisciplinary studies on lexical semantic divergences across time, domains, registers, and further language varieties.

This workshop presents research contributions across languages and across research disciplines to provide and compare resources, corpus-based empirical evidence and computational models for divergences in word meanings across language varieties. Relevant aspects include

- investigations on word sense definition and discrimination;
- corpus-based examples and discussions of lexical sense divergences;
- frequency distributions of word senses across corpora for language varieties;
- computational models to determine and measure lexical semantic change and divergence;
- relevance of word sense divergences for theories and applications in different fields.

References: Charles Boberg, John Nerbonne & Dominic Watt (eds.) (2018). The Handbook of Dialectology. Wiley-Blackwell. Marco Del Tredici & Raquel Fernandez (2017). Semantic variation in online communities of practice. In Proceedings of the 12th International Conference on Computational Semantics. Alessio Ferrari, Beatrice Donati, & Stefania Gnesi (2017). Detecting domain-specific ambiguities: An NLP approach based on wikipedia crawling and word embeddings. In Proceedings of the IEEE 25th International Requirements Engineering Conference Workshops. Anna Hätty, Dominik Schlechtweg & Sabine Schulte im Walde (2019). SURel: A gold standard for incorporating meaning shifts into term extraction. In Proceedings of the 8th Joint Conference on Lexical and Computational Semantics. Dirk Hovy & Christoph Purschke (2018). Capturing regional variation with distributed place representations and geographic retrofitting. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Hermann Paul (2002). Deutsches Wörterbuch: Bedeutungsgeschichte und Aufbau unseres Wortschatzes, 10th edition. Tübingen: Niemeyer. Dominik Schlechtweg, Anna Hätty, Marco del Tredici, & Sabine Schulte im Walde (2019). A wind of change: Detecting and evaluating lexical semantic change across times and domains. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy. ACL, 732–746. Dominik Schlechtweg, Sabine Schulte im Walde & Stefanie Eckmann (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. Martijn Wieling & John Nerbonne (2015). Advances in dialectometry. Annual Review of Linguistics 1, 243–264. Heike Wiese & Maria Pohle (2016). „Ich geh Kino" oder „... ins Kino"? Gebrauchsrestriktionen nichtkanonischer Lokalangaben. Zeitschrift für Sprachwissenschaft 35(2), 171–216.

AG 14