University of Stuttgart
Germany

DURel Annotation Tool

Human and Computational Measurement of Semantic Proximity, Sense Clusters and Semantic Change

March 26, 2025

Dominik Schlechtweg

Institute for Natural Language Processing, University of Stuttgart

Outline

Motivation

Background

Annotation Interface

Applications

Conclusion

Future work

References

Motivation

- **common problem**:
  - **given**: set of word uses (corpus)
  - **searched**: their meanings and their relations
- **relevant for**:
  - historical linguistics
  - lexicography
  - digital humanities
- **common approach**: researcher scans corpus himself
  - tedious
  - subjective
  - no protocol
  - bias
- **solution**: DURel Annotation Tool[1]
  - online interface
  - upload word uses for annotation
  - well-established protocol for contextual word meaning annotation

(Erk et al., 2013; Schlechtweg et al., 2020)

[1] `https://durel.ims.uni-stuttgart.de/`

Background

Data

A	1824	and taking a knife from her pocket, she opened a vein in her little **arm**,
B	1842	And those who remained at home had been heavily taxed to pay for the **arms**, ammunition;
C	1860	and though he saw her within reach of his **arm**, yet the light of her eyes seemed as far off
		...
D	1953	overlooking an **arm** of the sea which, at low tide, was a black and stinking mud-flat
E	1975	twelve miles of coastline lies in the southwest on the Gulf of Aqaba, an **arm** of the Red Sea.
F	1985	when the disembodied **arm** of the Statue of Liberty jets spectacularly out of the

Table 1: Sample of diachronic corpus.

Annotation

(A) [. . .] and taking a knife from her pocket, she opened a vein in her little **arm**, and dipping a feather in the blood, wrote something on a piece of white cloth, which was spread before her.

(D) It stood behind a high brick wall, its back windows overlooking an **arm** of the sea which, at low tide, was a black and stinking mud-flat [. . .]

Scale

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

Table 2: DURel relatedness scale.

Graph representation

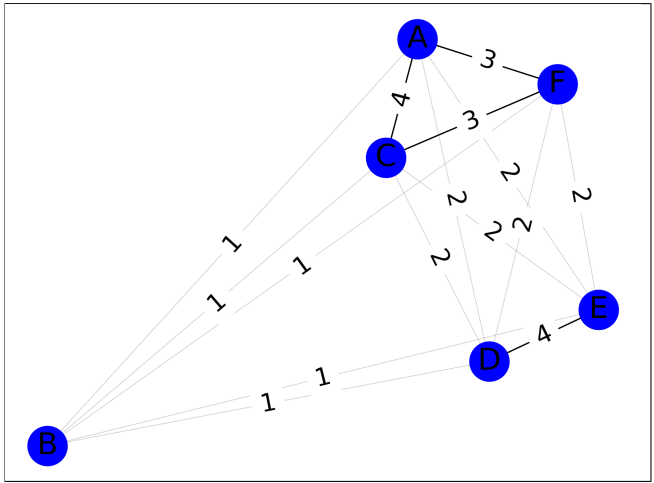


Figure 1: Word Usage Graph of English *arm*.

Clustering

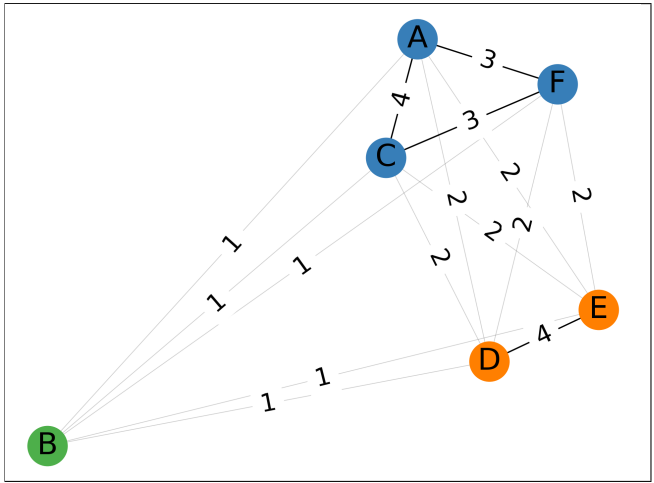
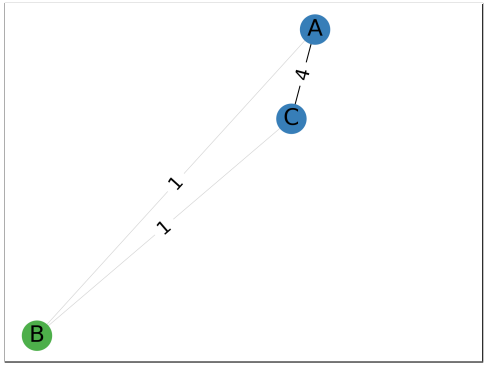
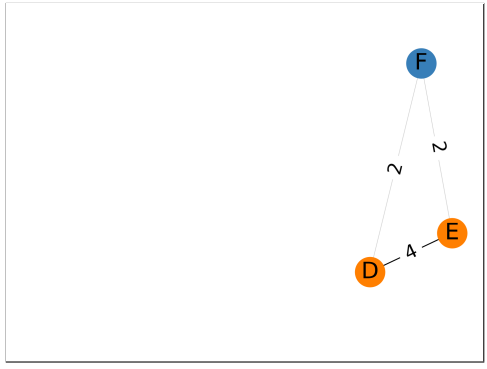


Figure 2: Word Usage Graph of English *arm*.

Lexical Semantic Change



t_1



t_2

Annotation Interface

Features

- **DURel annotation tool** (Schlechtweg, Virk, et al., 2024)
- data upload
  1. uses
  2. pairs
  3. judgments
- data inspection
- human annotation
- computational annotation
- annotator checking
- annotation statistics
- data clustering
- data visualization
- data export

Demo

- https://durel.ims.uni-stuttgart.de/

Applications

Case Study: Lexical Semantic Change Discovery

- **task**: (Kurtyigit et al., 2021)
  - discover novel word senses over time in a German diachronic corpus pair
- **use**:
  - post hoc annotate model predictions for validation
- **results**:
  - good predictions
  - reasonable agreement
  - useful visualizations
  - discovered unrecorded changes

Case Study: Lexical Semantic Change Discovery

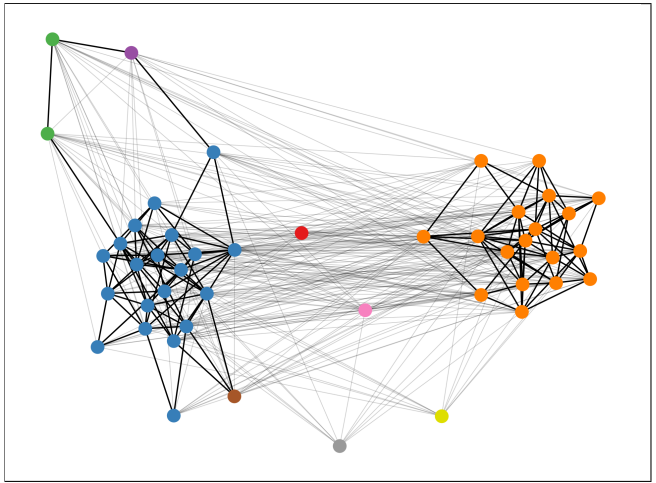
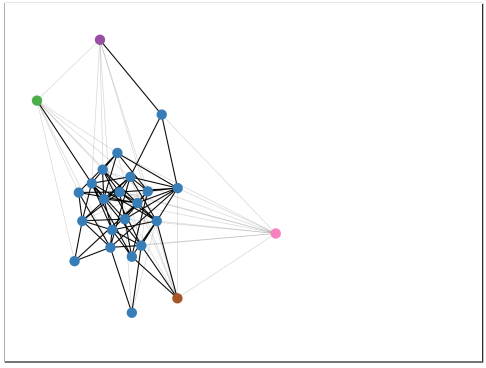
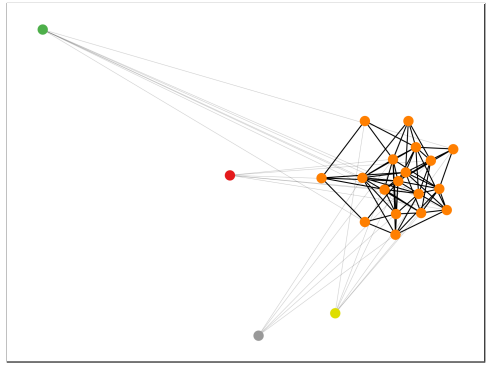


Figure 4: Word Usage Graph of German *Zehner*.

Case Study: Lexical Semantic Change Discovery



t_1



t_2

Case study: Polysemy/Synonymy across Language Varieties

- **task**: (Baldissin et al., 2022)
  - detect diatopic lexical semantic variation in Spanish
- **use**:
  - annotate data samples from Spanish variants
  - extend DURel framework to onomasiological questions
- **results**:
  - reasonable agreement
  - nice graphs
  - annotation across lemmas works well

Case study: Polysemy/Synonymy across Language Varieties

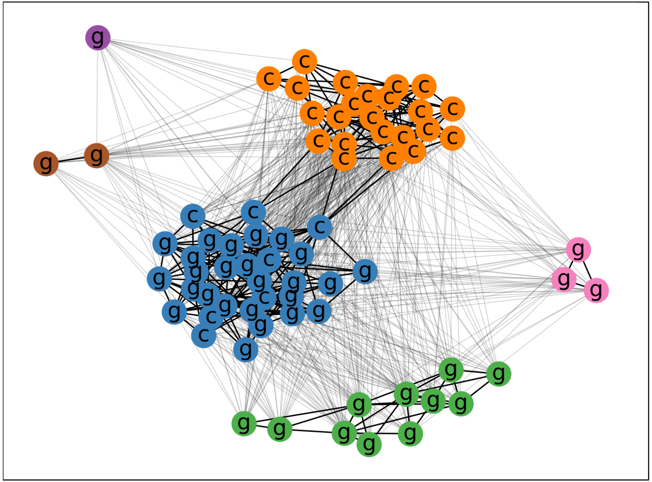
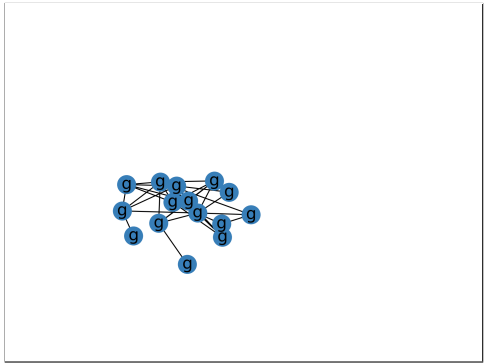
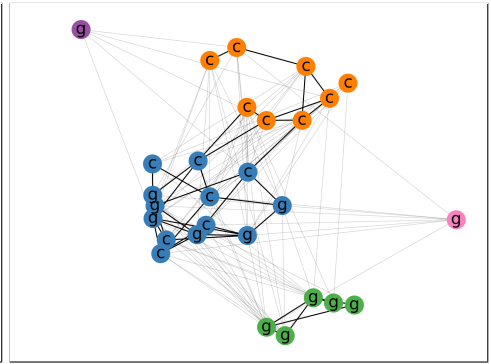


Figure 6: Word Usage Graph of Spanish *colectivo* and *guagua*.

Case study: Polysemy/Synonymy across Language Varieties



Cuba



Argentina

Case study: Automated Clustering for Dictionary Updating

- **task**: (Sander et al., 2024; Sköldberg et al., 2024)
  - reduce the workload when searching for new dictionary headword senses
- **use**:
  - cluster uses for headwords based on automatic pairwise annotations
  - compare cluster number to sense entry number in dictionary
- **results**:
  - works well
  - discovered metonymic & metaphoric uses
  - discovered proper nouns
  - scalable

Case study: Automated Clustering for Dictionary Updating

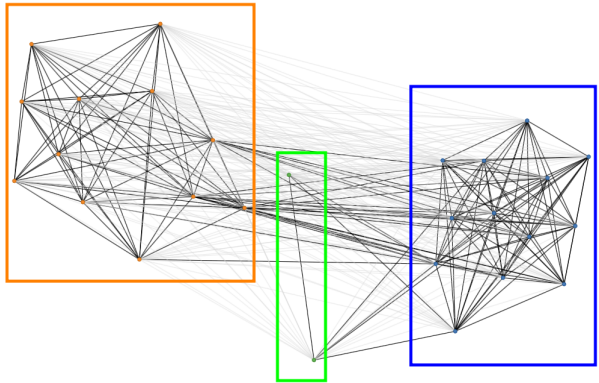


Figure 8: Usage clustering of *to seep* in three clusters. From left to right: metaphorical use, PoS-errors, literal use.

Studies

- post-hoc annotation of **German** semantic changes (Kurtyigit et al., 2021)
- **Spanish** diatopic variation dataset (Baldissin et al., 2022)
- **Spanish** LSCD dataset (Zamora-Reina et al., 2022)
- **Norwegian** LSCD dataset (Kutuzov et al., 2022)
- **Russian** WSI dataset (Aksenova et al., 2022)
- **Chinese** LSCD dataset (Chen et al., 2023)
- extended LSCD datasets for **English**/**German**/**Swedish** (Schlechtweg, Cassotti, et al., 2024)
- WSI with automatic annotators for **English**/**German**/**Swedish** (Sander et al., 2024; Sköldberg et al., 2024)
- manual WSI for **German** (Theuer Linke, 2023)
- **Dutch** LSCD data set (ongoing)
- ...

Conclusion

- **inter-subjectivity**:
  - avoids experimenter bias through standard protocol and annotation by multiple humans
  - inter-annotator agreement gives measure of reliability
- **simple**:
  - the judgment of use pair relatedness is an intuitive task for annotators generally yielding high agreement (Erk et al., 2013; Schlechtweg et al., 2018)
  - annotated data can be visualized as semantic relatedness graphs on 2D plots
- **preparation-lean**:
  - researchers only need to sample word uses
- **grounded in theory**:
  - relatedness judgments have theoretical basis in cognitive semantics (Blank, 1997; Schlechtweg et al., 2018)
- **flexible**:
  - clustering algorithm and parameters can be changed after annotation, avoiding re-annotation
- **automatatable**:
  - pairwise judgments can be done automatically for large sets of data with good performance

Future work

- optimized sampling
- annotation history + modifying previous annotations
- optimization for large-scale automated prediction
- corpus interface
- **what do you need?**
- **funding**

References I

Aksenova, A., Gavrishina, E., Rykov, E., & Kutuzov, A. (2022, October). RuDSI: Graph-based word sense induction dataset for Russian. In D. Ustalov et al. (Eds.), *Proceedings of textgraphs-16: Graph-based methods for natural language processing* (pp. 77–88). Gyeongju, Republic of Korea: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2022.textgraphs-1.9

Baldissin, G., Schlechtweg, D., & Schulte im Walde, S. (2022). DiaWUG: A Dataset for Diatopic Lexical Semantic Variation in Spanish. In *Proceedings of the 13th language resources and evaluation conference*. Marseille, France: European Language Resources Association. Retrieved from https://aclanthology.org/2022.lrec-1.278/

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

Chen, J., Chersoni, E., Schlechtweg, D., Prokic, J., & Huang, C.-R. (2023). ChiWUG: A graph-based evaluation dataset for Chinese lexical semantic change detection. In *Proceedings of the 4th international workshop on computational approaches to historical language change*. Singapore: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2023.lchange-1.10/

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

Kurtyigit, S., Park, M., Schlechtweg, D., Kuhn, J., & Schulte im Walde, S. (2021, aug). Lexical Semantic Change Discovery. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)*. Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2021.acl-long.543/

Kutuzov, A., Touileb, S., Mæhlum, P., Enstad, T., & Wittemann, A. (2022, June). NorDiaChange: Diachronic semantic change dataset for Norwegian. In *Proceedings of the thirteenth language resources and evaluation conference* (pp. 2563–2572). Marseille, France: European Language Resources Association. Retrieved from https://aclanthology.org/2022.lrec-1.274

Sander, P., Hengchen, S., Zhao, W., Ma, X., Sköldberg, E., Virk, S. M., & Schlechtweg, D. (2024). The DURel Annotation Tool: Using fine-tuned LLMs to discover non-recorded senses in multiple languages. In *Proceedings of the Workshop on Large Language Models and Lexicography at 21st EURALEX International Congress Lexicography and Semantics*. Retrieved from https://www.cjvt.si/wp-content/uploads/2024/10/LLM-Lex_2024_Book-of-Abstracts.pdf

Schlechtweg, D., Cassotti, P., Noble, B., Alfter, D., Schulte Im Walde, S., & Tahmasebi, N. (2024, nov). More DWUGs: Extending and evaluating word usage graph datasets in multiple languages. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 conference on empirical methods in natural language processing* (pp. 14379–14393). Miami, Florida, USA: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2024.emnlp-main.796

References II

Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., & Tahmasebi, N. (2020). SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. In *Proceedings of the 14th International Workshop on Semantic Evaluation*. Barcelona, Spain: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/2020.semeval-1.1/

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. Retrieved from https://www.aclweb.org/anthology/N18-2027/

Schlechtweg, D., Virk, S. M., Sander, P., Sköldberg, E., Theuer Linke, L., Zhang, T., ... Schulte im Walde, S. (2024, mar). The DURel annotation tool: Human and computational measurement of semantic proximity, sense clusters and semantic change. In N. Aletras & O. De Clercq (Eds.), *Proceedings of the 18th conference of the european chapter of the association for computational linguistics: System demonstrations* (pp. 137–149). St. Julians, Malta: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2024.eacl-demo.15

Sköldberg, E., Virk, S. M., Sander, P., Hengchen, S., & Schlechtweg, D. (2024). Revealing semantic variation in Swedish using computational models of semantic proximity: Results from lexicographical experiments. In *Proceedings of the 21st EURALEX International Congress Lexicography and Semantics*. Retrieved from https://euralex.org/publications/revealing-semantic-variation-in-swedish-using-computational-models-of-semantic-proximity-results-from-lexicographical-experiments/

Theuer Linke, L. (2023). *Testing the effect of using crowdsourced semantic proximity judgments in the process of human lexicographical word sense clustering* (Bachelor thesis). University of Stuttgart.

Zamora-Reina, F. D., Bravo-Marquez, F., & Schlechtweg, D. (2022). LSCDiscovery: A shared task on semantic change discovery and detection in Spanish. In *Proceedings of the 3rd international workshop on computational approaches to historical language change*. Dublin, Ireland: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2022.lchange-1.16/