



# DURel Annotation Tool

## Prospects on a workbench for lexicographers

Maïke Park Dominik Schlechtweg

### Introduction

#### What is DURel?

An online annotation interface using **proximity judgments of use pairs from human annotators** to infer cluster structure **consecutively** and **efficiently**.

#### Why use it for lexicographic purposes?

- ▶ **objective:**
  - ▶ avoids 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:** lexicographers 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

### Overview of the system

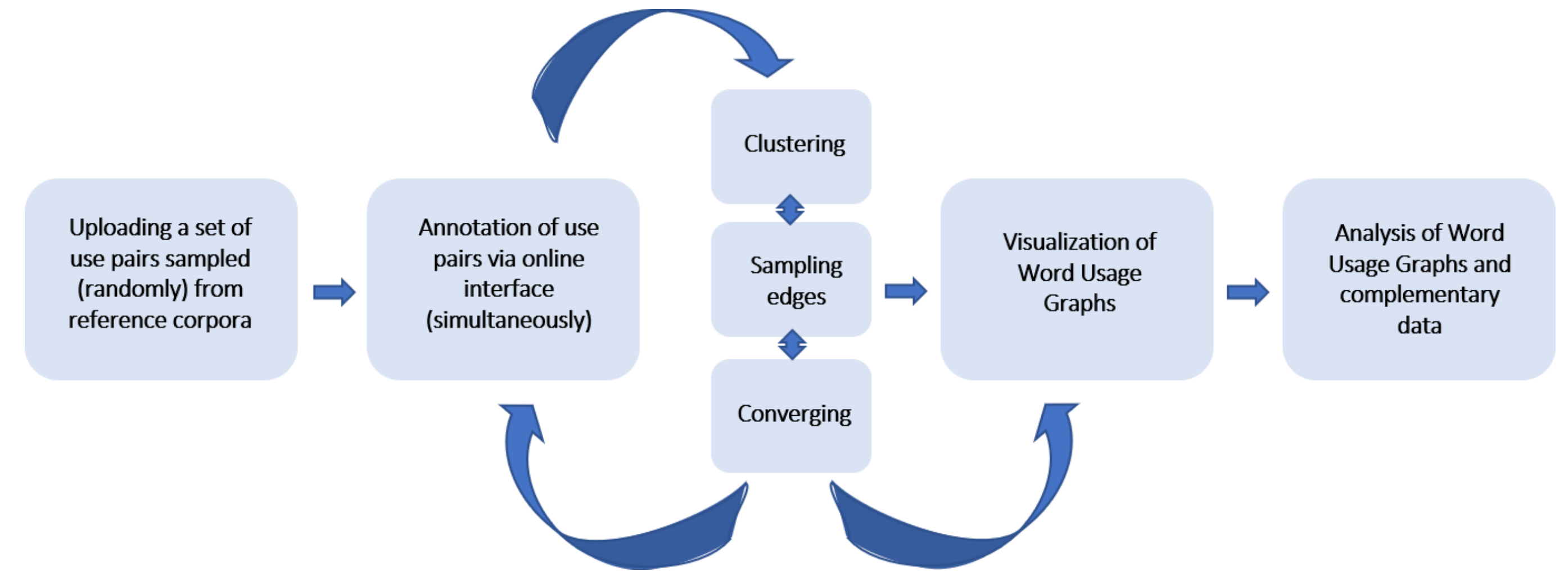
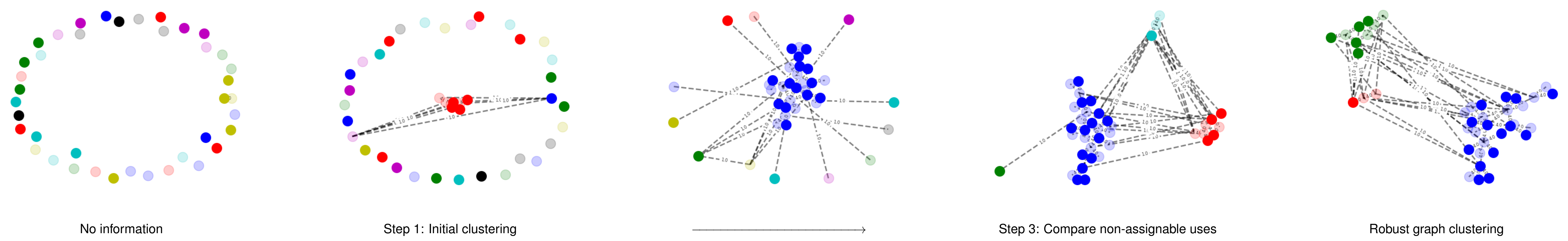


Figure: Prospective workflow using DURel's online interface.

### Annotation Progress



### Online interface

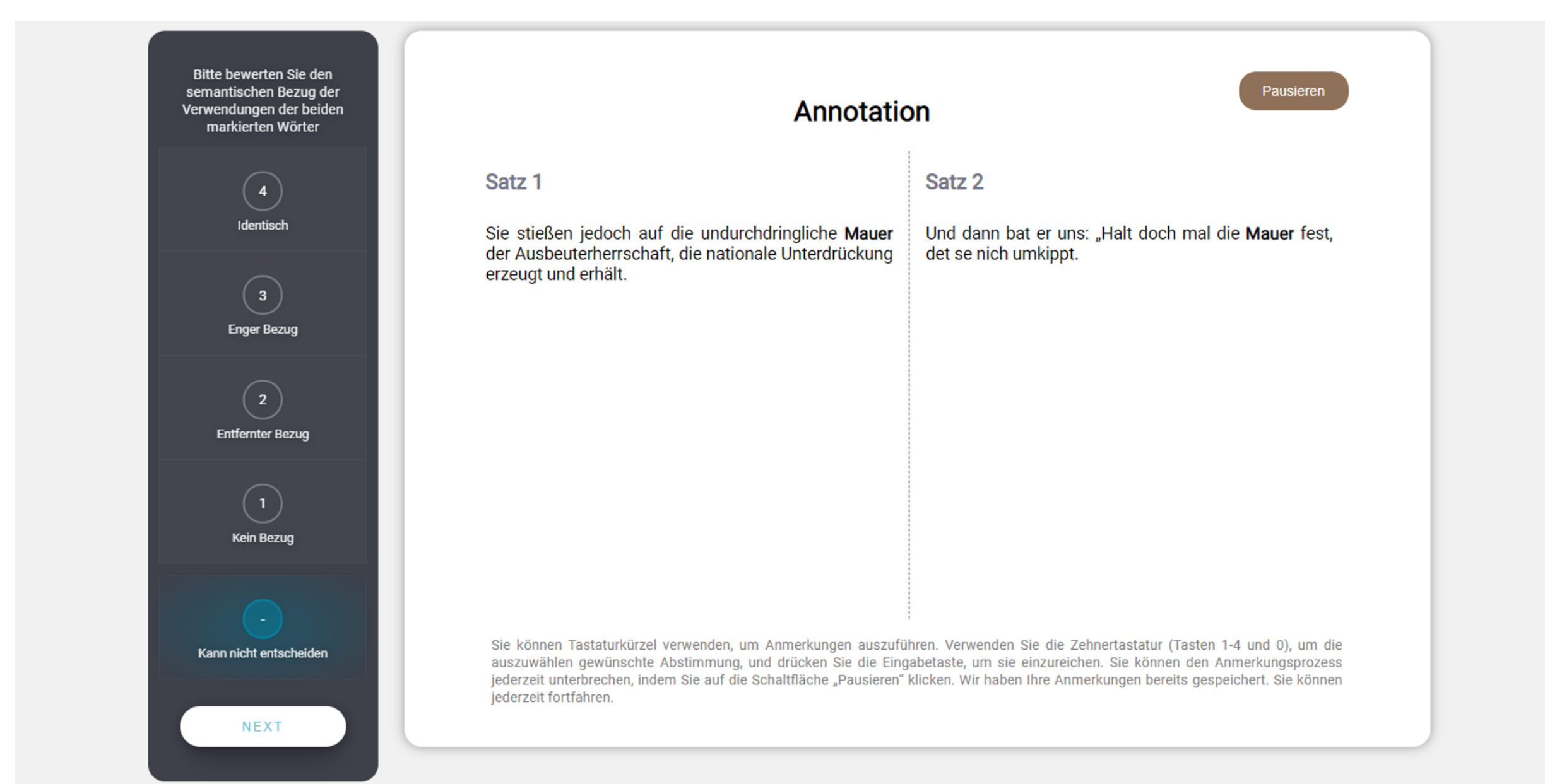


Figure: Screenshot of annotation interface with 4-point scale of relatedness [Erk et al. 2013, Schlechtweg et al. 2018], German version.

### Visualizing results

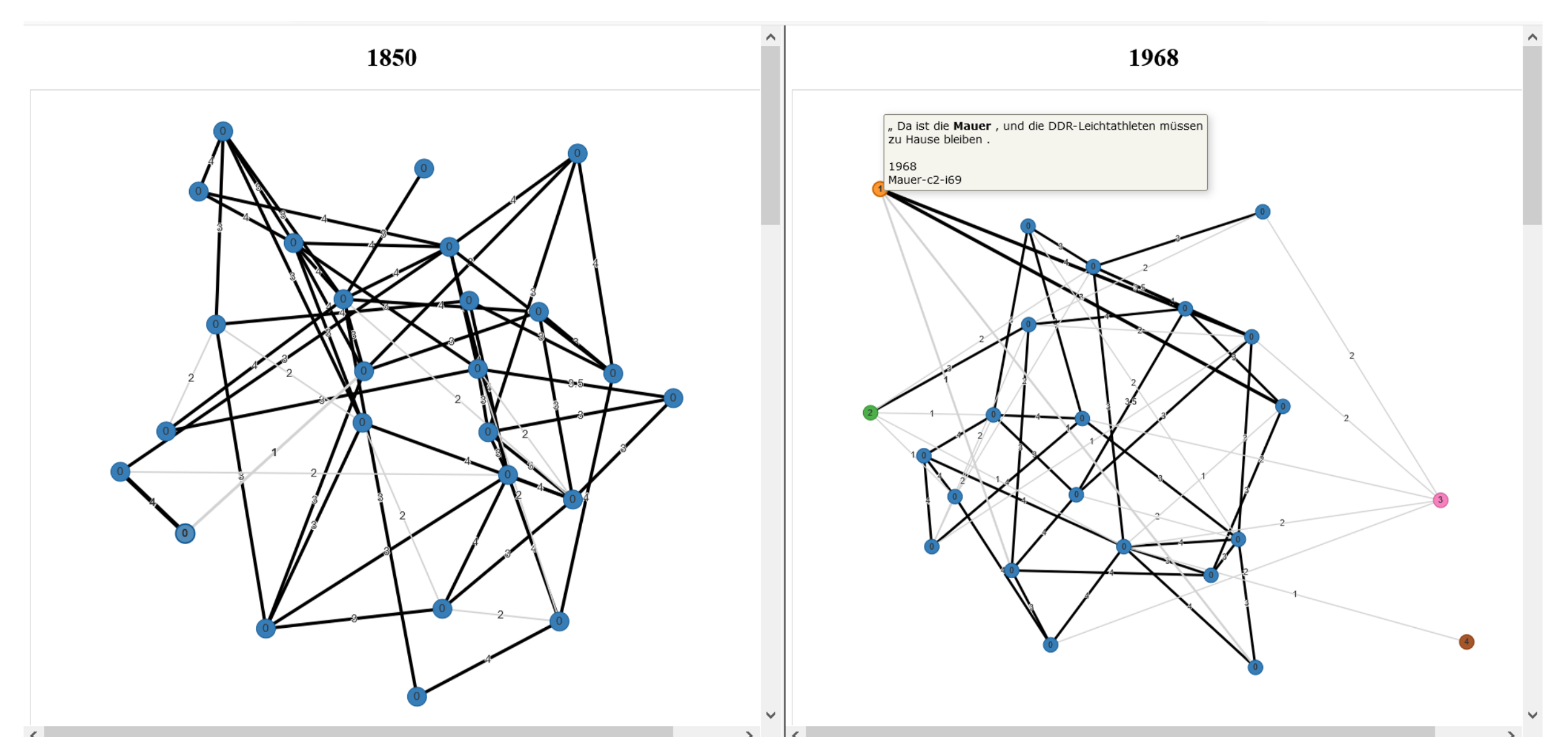


Figure: Word usage graphs for *Mauer* ('wall') displaying lexical semantic change over time between  $t_1$  and  $t_2$ , with meta-information exhibited for one node on the right.

### Applications

#### First test run with lexicographic analysis:

- ▶ discovery of actually novel senses in diachronic corpora
- ▶ confirmation of lexical semantic change over time (as included in dictionaries) [cf. Kurtyigit et al. 2021]

#### Upcoming studies:

- ▶ measuring semantic patterns of infrequent semantic neologisms from corpora representing Contemporary German compiled from newspaper data

### Outlook

- ▶ further development of the online web interface
- ▶ improving algorithms for the discovery of novel lexical semantic change
- ▶ exploring citizen science approaches to acquire (steady) support for annotations

#### Acknowledgments

The first author was supported by the Konrad Adenauer Foundation and the CRETA center funded by the German Ministry for Education and Research (BMBWF) during the conduct of this research. The system's beta version was implemented by Annalena Streichert, Anne Reuter, Enrique Waldo Medina Castaneda and Lukas Theuer Linke.

#### References

- Blank, A. (1997). *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*. Tübingen: Niemeyer.
- 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). Lexical Semantic Change Discovery. In *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Online: Association for Computational Linguistics.
- 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/>