

# The DUREL Annotation Tool: Human and Computational Measurement of Semantic Proximity, Sense Clusters and Semantic Change

Dominik Schlechtweg,<sup>✦</sup> Shafqat Mumtaz Virk,<sup>✦</sup> Pauline Sander,<sup>✦</sup> Emma Sköldberg,<sup>✦</sup> Lukas Theuer Linke,<sup>✦</sup>  
Tuo Zhang,<sup>✦</sup> Nina Tahmasebi,<sup>✦</sup> Jonas Kuhn,<sup>✦</sup> Sabine Schulte im Walde<sup>✦</sup>

<sup>✦</sup>University of Stuttgart, <sup>✦</sup>University of Gothenburg

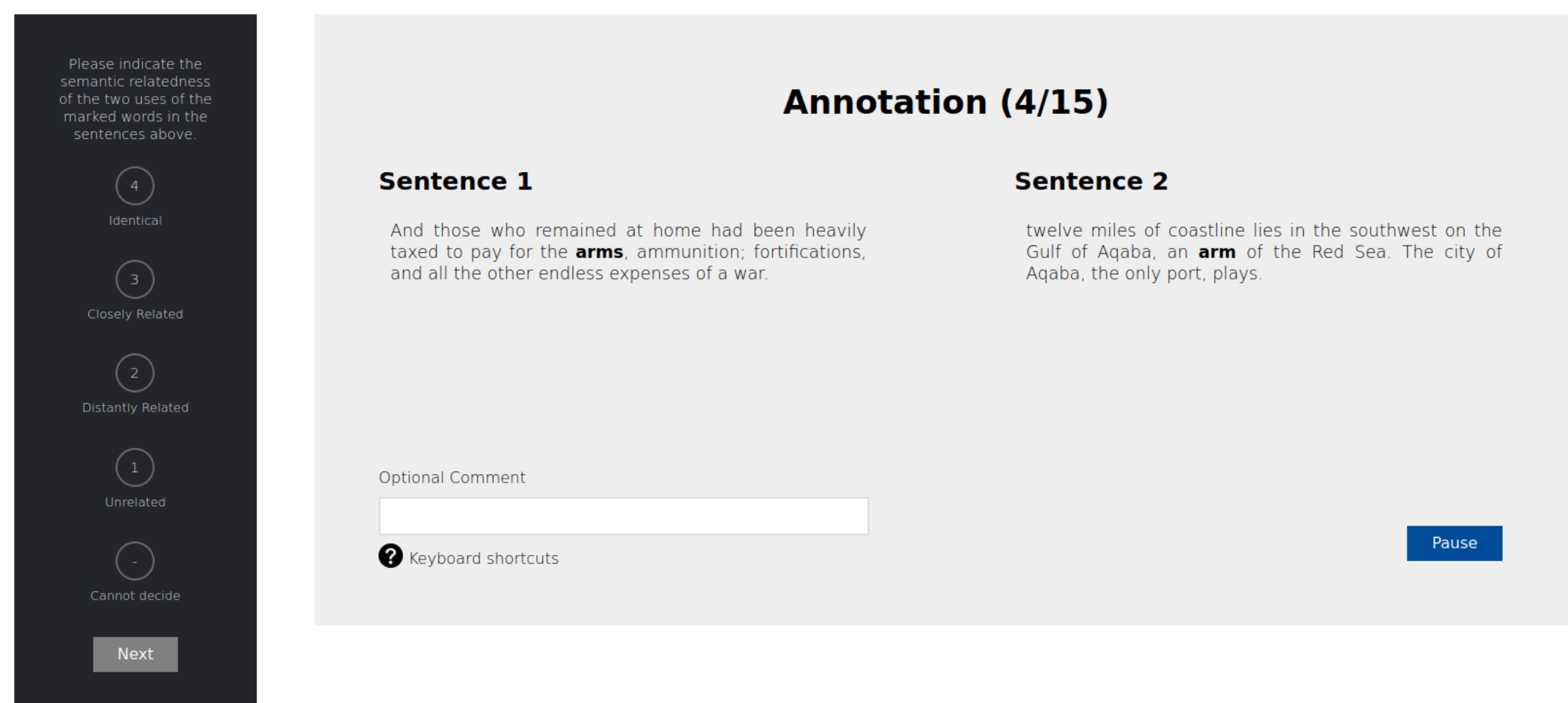
## Semantic Proximity

- ▶ How much do the meanings of two word uses “have in common”?
- ▶ can be measured using graded judgments of word usage pairs  
[Erk et al. 2013, Schlechtweg et al. 2018]
- ▶ judgments are clustered and populate weighted graphs which make senses visible  
[Bansal et al. 2004, Schlechtweg et al. 2020]
- ▶ essential for **lexicographic clustering** and distinction of **word senses**
  - ▶ but difficult to measure and large workload

## Tool Overview

- ▶ **data inspection** and automatic **data validation** for researchers,
- ▶ **guidelines** in multiple languages and **tutorials** for annotator training,
- ▶ an intuitive task interface for annotators,
- ▶ **computational annotators** for the annotation of large scale projects,
- ▶ basic **statistical analysis**, and
- ▶ graph **clustering and visualization** in WUGs (Word Usage Graphs)  
<https://durel.ims.uni-stuttgart.de>

## Human Annotation



**Annotation (4/15)**

**Sentence 1**  
And those who remained at home had been heavily taxed to pay for the **arms**, ammunition, fortifications, and all the other endless expenses of a war.

**Sentence 2**  
twelve miles of coastline lies in the southwest on the Gulf of Aqaba, an **arm** of the Red Sea. The city of Aqaba, the only port, plays.

Optional Comment: \_\_\_\_\_

Keyboard shortcuts: \_\_\_\_\_

Pause

Human annotation window. Scale adapted from Schlechtweg et al. [2018].

## Data Inspection

**Uses** Back

ID	Lemma	POS	Date	Left Context	Target	Right Context
1233516	arm	N	1824	and taking a knife from her pocket, she opened a vein in her little	arm	, and digging a feather in the blood, wrote something on a piece of white cloth, which was spread before her.
1233517	arm	N	1842	And those who remained at home had been heavily taxed to pay for the	arms	, ammunition, fortifications, and all the other endless expenses of a war.
1233518	arm	N	1860	and though he saw her within reach of his	arm	, yet the light of her eyes seemed as far off as that of a
1233519	arm	N	1953	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
1233520	arm	N	1975	twelve miles of coastline lies in the southwest on the Gulf of Aqaba, an	arm	of the Red Sea. The city of Aqaba, the only port, plays.
1233521	arm	N	1985	when the disembodied	arm	of the Statue of Liberty jets spectacularly out of the sandy beach.

Project with uses of *arm* sampled from the COHA corpus [Davies 2012].

## Data Analysis

- ▶ provided summary statistics include
  - ▶ various measures of **annotator agreement**
  - ▶ **label counts and averages** for words, groupings and annotators
  - ▶ semantic variation measures and **semantic change measures**

Example Pairs Show

Example Uses

Lemma	#Uses	#Pairs	#Ann	Ann	#Judg	#0	#1	#2	#3	#4	Avg.
arm	6	15	7	Random, Ipvovarovova, XL-Lexeme-Cosine, PaulineSander, billnoble, XL-Lexeme-Binary, Jessica	76	0	26	9	9	17	1.8289474
All Words	6	15	7	Random, Ipvovarovova, XL-Lexeme-Cosine, PaulineSander, billnoble, XL-Lexeme-Binary, Jessica	76	0	26	9	9	17	1.8289474

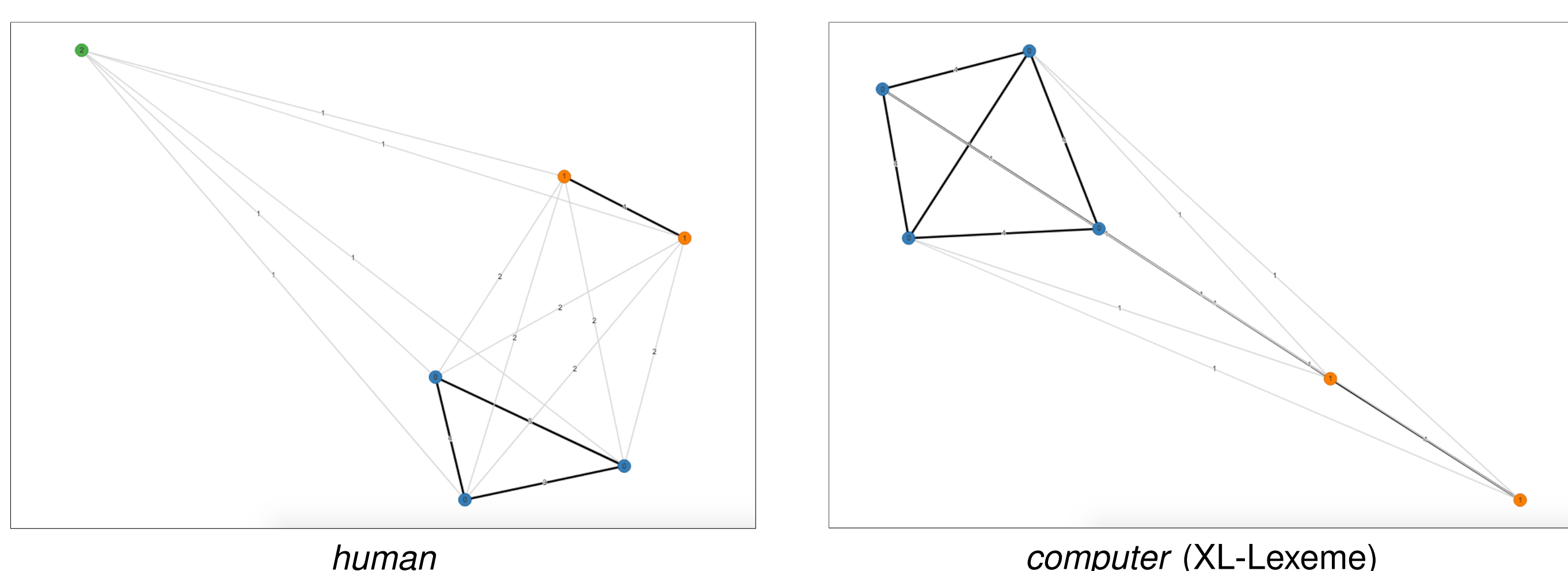
Simple database statistics.

## Computational Annotation

- ▶ researchers can select multi-lingual Word-in-Context (WiC) models as annotators
  - ▶ **Random** samples a random integer between 1 and 4 (as baseline)
  - ▶ **XLMR** [Conneau et al. 2019] vectorizer with multi-layer perceptron and binary classification (1 or 4)
  - ▶ **XL-Lexeme** bi-encoder that uses a XLMR-based Siamese Network [Cassotti et al. 2023]; trained to minimize the contrastive loss with cosine distance on several WiC datasets; predicts either value 1 or 4 based on thresholding

## Case (i): The *arm* Example

- ▶ computational annotation with XL-Lexeme
- ▶ all annotations clustered using correlation algorithm



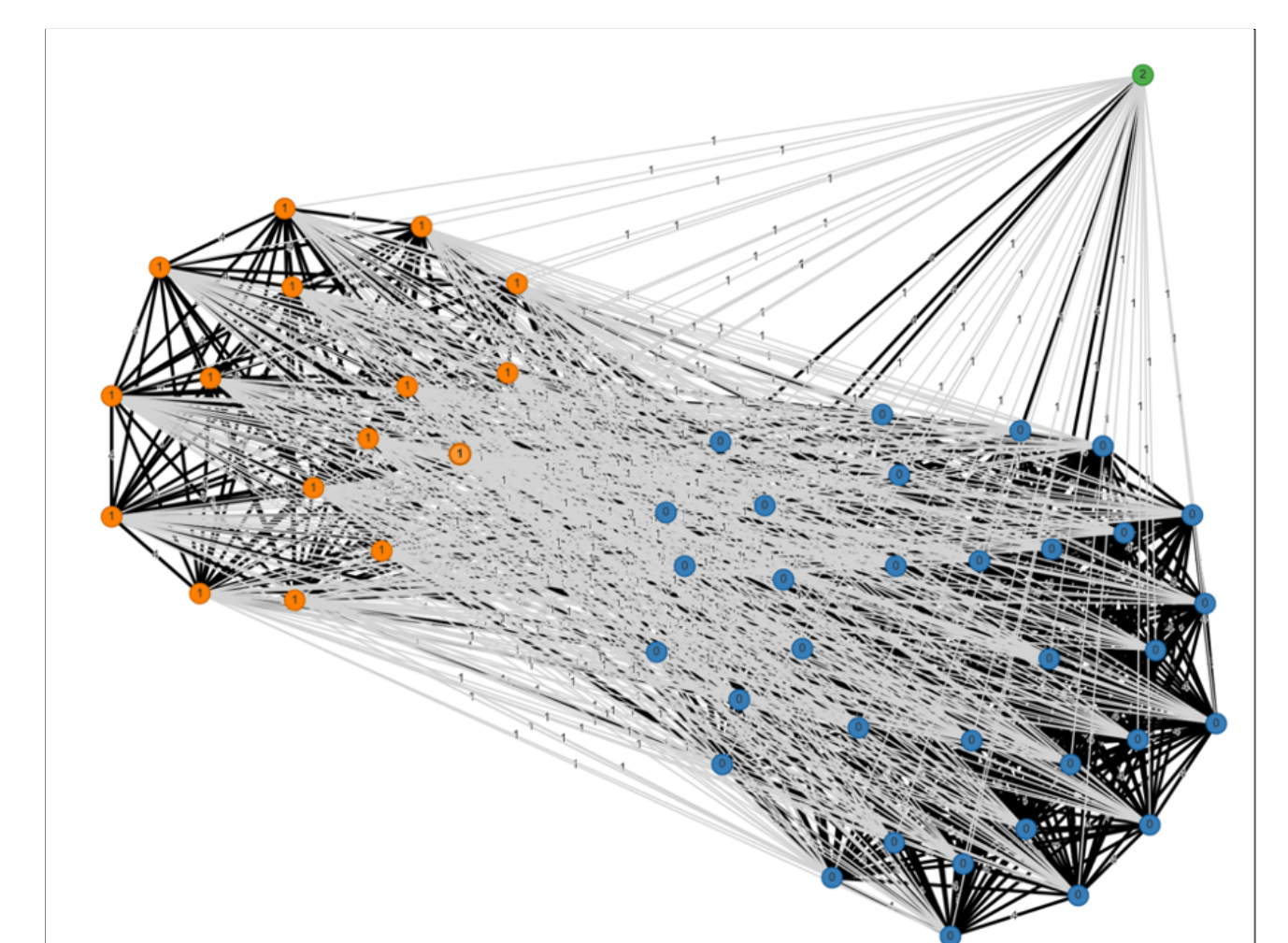
- ▶ orange cluster corresponds to metaphorical sense ‘arm of the sea’ in both cases; computational annotator merges ‘body part’ and ‘weapon’ sense

## Case (ii): Lexicographer Study

- ▶ revealing semantic variation in Swedish
  - ▶ related to the revision of ‘The Contemporary Dictionary of the Swedish Academy’  
<https://svenska.se>  
[Borin et al. 2012]
  - ▶ 18 words, 50 sentences each from the SVT corpus
  - ▶ automatically paired and annotated with XL-Lexeme
  - ▶ clustered using the correlation algorithm

Word	ARI	Word	ARI
ofantlig	1.0	klimat	0.083
enkelspårig	1.0	vansinnig	0.0
baksida	0.912	lirka	0.0
bagage	0.785	kapitulera	0.0
fasad	0.652	hemmaplan	0.0
vissen	0.645	hagla	0.0
skör	0.507	fotoavtryck	0.0
rutten	0.333	tvärnita	-0.019
ventilera	0.303	kriga	-0.025
<b>Average</b>		<b>0.343</b>	

Cluster evaluation based on ARI.



## Acknowledgments

The development of the tool has been funded by the project *Towards Computational Lexical Semantic Change Detection* supported by the Swedish Research Council (2019–2022; contract 2018-01184), the research program *Change is Key!* supported by Riksbankens Jubileumsfond (under reference number M21-0021) and the CRETA center funded by the German Ministry for Education and Research (BMBF). We thank Anne Reuter, Annalena Streichert, Enrique Medina Waldo Castaneda, Sinan Kurtyigit, Serge Kotchourko, Pedro G. Bascosy, Francesco Periti and Jing Chen for contributions to this project.

## References

- Bansal, N., Blum, A., & Chawla, S. (2004). Correlation clustering. *Machine Learning*, 56(1-3), 89–113. doi: 10.1023/B:MACH.0000033116.57574.95
- Borin, L., Forsberg, M., & Roxendal, J. (2012, May). Korp — the corpus infrastructure of Språkbanken. In *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)* (pp. 474–478). Istanbul, Turkey: European Language Resources Association (ELRA). Retrieved from <http://www.lrec-conf.org/proceedings/lrec2012/pdf/248.Paper.pdf>
- Cassotti, P., Siciliani, L., de Gemmis, M., Semeraro, G., & Basile, P. (2023, July). Xl-lexeme: Wic pretrained model for cross-lingual lexical semantic change. In *Proceedings of the 61th annual meeting of the association for computational linguistics*. Online: Association for Computational Linguistics.
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.
- Davies, M. (2012). Expanding Horizons in Historical Linguistics with the 400-Million Word Corpus of Historical American English. *Corpora*, 7(2), 121–157.
- Erk, K., McCarthy, D., & Gaylord, N. (2013). Measuring word meaning in context. *Computational Linguistics*, 39(3), 511–554.
- 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/>