University of Stuttgart
Germany

CHANGE
IS KEY

The DURel Annotation Tool

Using fine-tuned LLMs to discover non-recorded senses in multiple languages

October 16, 2024

Pauline Sander, Simon Hengchen, Wei Zhao, Xiaocheng Ma, Emma Sköldberg, Shafqat Virk and Dominik Schlechtweg

LLM-LEX Workshop 2024

IGUANODON AI

UNIVERSITÉ
DE GENÈVE

UNIVERSITY OF GOTHENBURG

Introduction

- **Goal:** Reduce the workload when searching for new headword senses.

  ↪ DURel Annotation Tool
  - automatic semantic proximity annotation
  - graph clustering and visualization
  - guide lexicographers in their work

Semantic Proximity Scale

4: Identical	Identity
↑ 3: Closely Related	↑ Context Variance
2: Distantly Related	Polysemy
1: Unrelated	Homonymy

Table 1: The DURel relatedness scale (Schlechtweg et al., 2018) on the left and its interpretation from Schlechtweg (2023, p. 33) on the right.

Word-in-Context Models

- produce contextualized embeddings
- trained on large amounts of data
- optimized on human-annotated semantic proximity judgments

XL-Lexeme (Cassotti et al., 2023)

- bi-encoder
- vectorizes the input sequences using a XLMR-based Siamese Network
- minimize the contrastive loss with cosine distance

  ↪ running within the DURel Tool

Arm Example

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

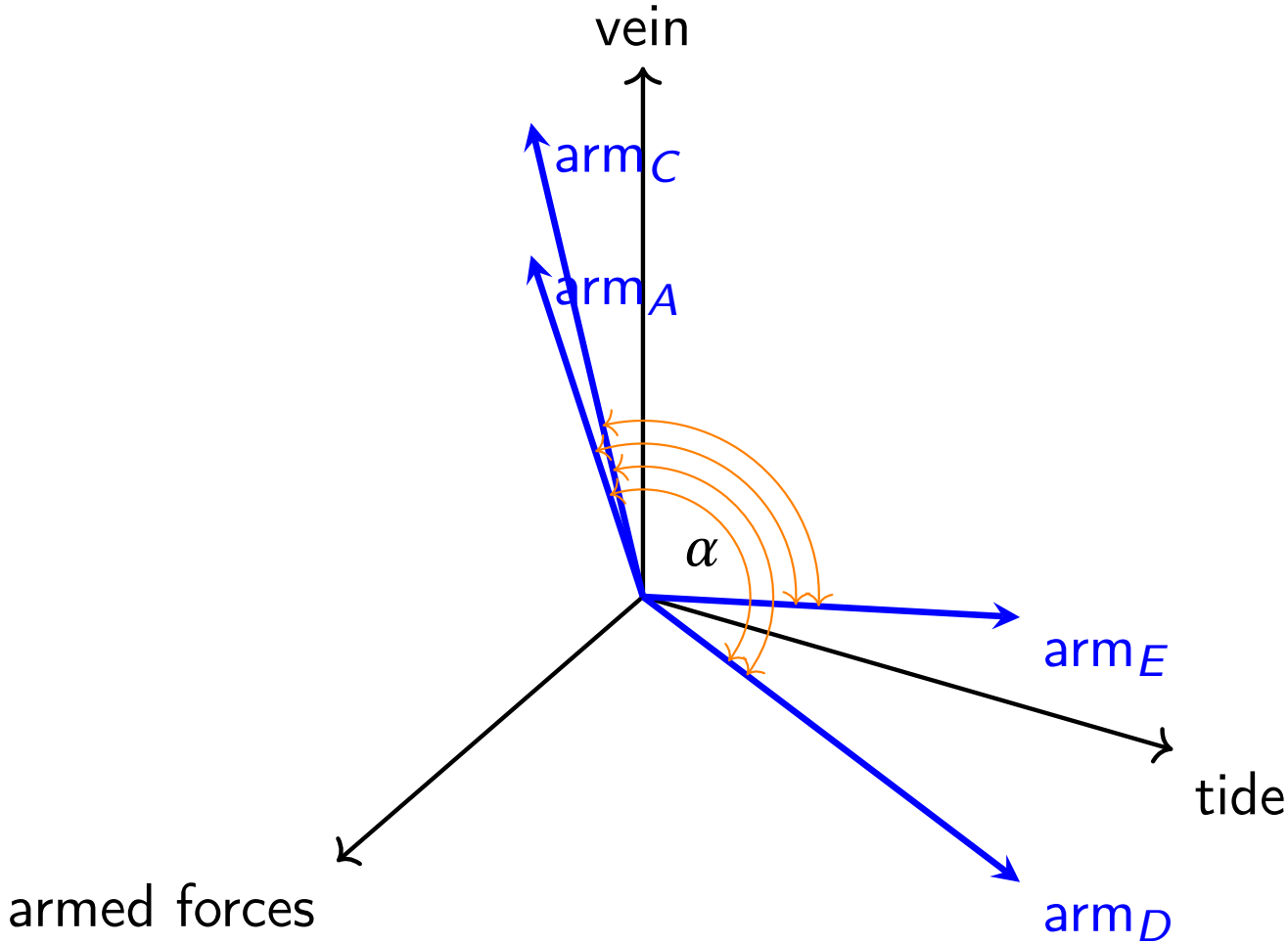
Table 2: Sample of diachronic corpus.

Word Use Pairs

(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 [. . .]

Automated Annotation of Semantic Proximity



Graph Representation

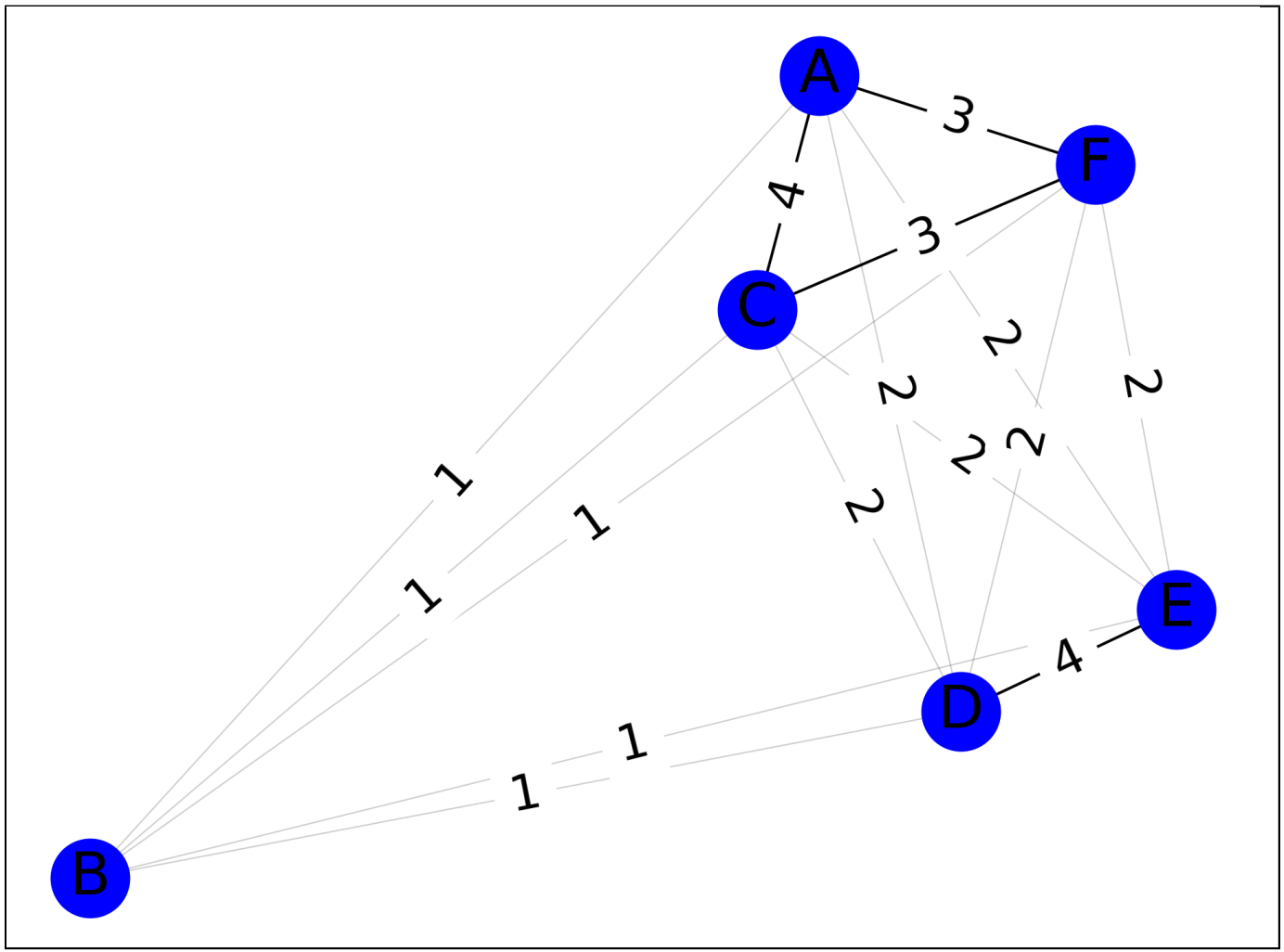


Figure 1: Word Usage Graph of English *arm*.

Clustering

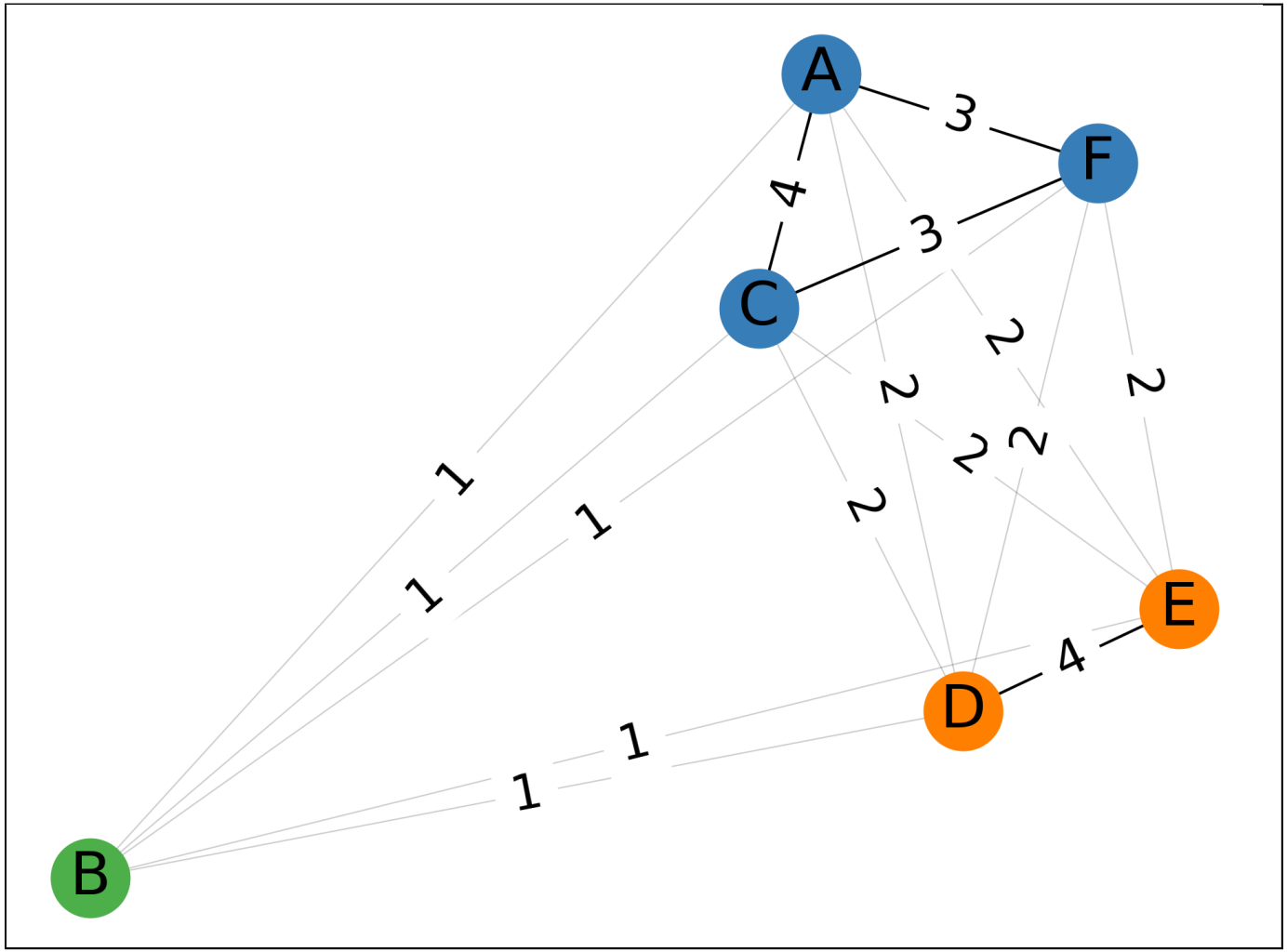


Figure 2: Word Usage Graph of English *arm*. $D = (3, 2, 1)$.

Application

Sense Discovery Task: Decide for a **large set of unseen** words which ones likely have gained new senses

(Sköldberg et al., 2024)

Sköldberg et al. (2024): Swedish

randomly select 281 headwords from SO

↓

sample 25 occurrences per headword from SVT corpus

↓

run through the DURel tool

↓

66 words (23%) predicted to have more than one cluster

↓

manual inspection

Our Experiments

Language	Dictionary	Corpus
English	WordNet[1]	2023 1M Leipzig corpus[2]
English	OED[3]	2015-2024 1M Leipzig corpus[4]
German	DWDS[5]	2023 1M Leipzig corpus[6]

Table 3: Overview of languages, dictionaries and corpora covered.

[1] WordNet (1998)
[2] Leipzig Corpora Collection (2023a)
[3] OED (2009)
[4] Leipzig Corpora Collection (2023a)
[5] DWDS (2024)
[6] Leipzig Corpora Collection (2023b)

Procedure: WordNet and DWDS

1. randomly sample headwords from the respective language's **Wiktionary**
2. look up the senses of all headwords in the dictionary
3. **select one PoS if multiple apply**
4. only keep monosemous headwords
5. sample 25 uses per headword from the **lemmatized** Leipzig corpus
6. run the DURel pipeline
7. manually inspect the results

Procedure: OED

1. sample headwords from **OED updates**[7] (2015-2024)
2. look up the senses of all headwords in the dictionary
3. -
4. only keep monosemous headwords
5. sample 25 uses per headword from the **unlemmatized** Leipzig corpus
6. run the DURel pipeline
7. manually inspect the results

[7] The differences to the other procedure are highlighted in bold.

Results

	Sampled	>1 cluster		Named entity		New sense	
WordNet	100	40	40%	15	37.5%	4	10%
OED	103	44	42.7%	20	45.4%	1	2.2%
German	108	46	42.6%	9	19.5%	8	17.3%

Table 4: Quantitative results for all experiments

Results: WordNet

- 40 words with more than one cluster
- 16 words with no meaningful sense distinction
- 5 words with faulty uses or clustering errors
- 15 words with named entities as second cluster
  - 'euphoria' the state of mind vs 'Euphoria' the TV show
  - 'Lisbon', Portugal vs 'Lisbon', Ohio (US)
- **4 words with new senses** [8]
  - 'water seeping' vs 'emotion seeping'
  - 'forest logging' vs 'computer logging'

[8] *New* in the context of the dictionary.

Example: 'seeping'

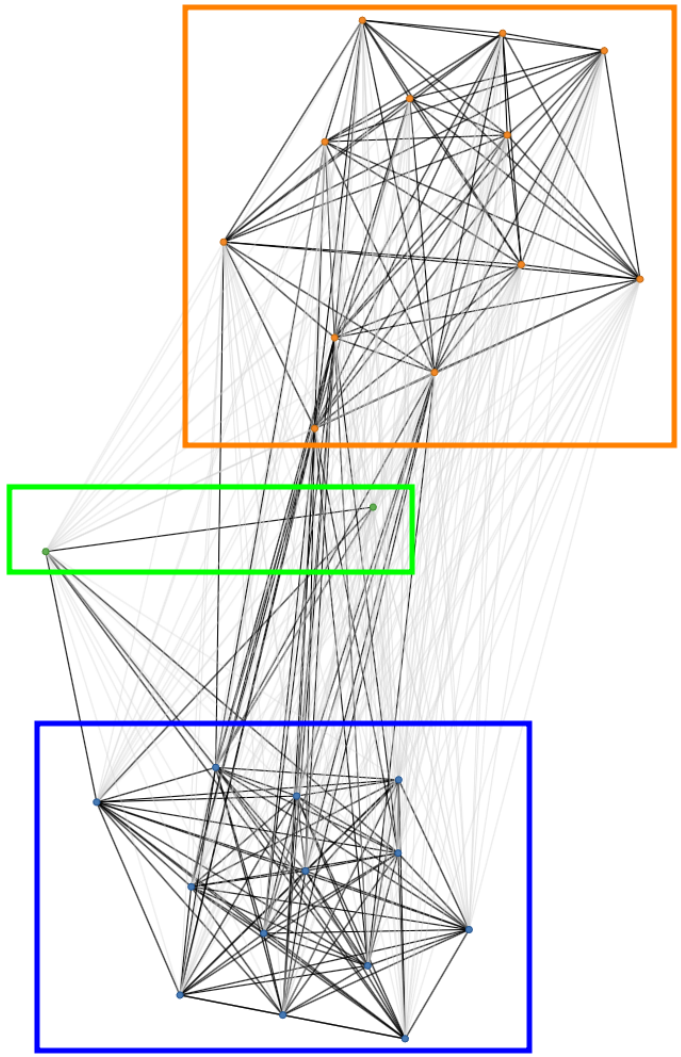


Figure 3: Usage clustering of 'to seep' in three clusters. From top to bottom: metaphorical use, PoS-errors, literal use.

Results: DWDS

- 46 words with more than one cluster
- 26 words with no meaningful sense distinction
- 3 words with faulty uses or clustering errors
- 9 words with named entities as second cluster
  - 'Alphabet' (the company vs 'alphabet')
- **8 words with new senses**
  - Einlassung' ('statement' vs 'mounting')
  - 'Einmarsch' ('march-in' vs 'invasion')
  - 'Segnung' ('the act of blessing' vs 'the blessing itself')
  - 'Eck' ('street corner' vs 'corner of a goal')
  - 'Lebenswerk' ('life's work' vs 'charitable organization)

Results: OED

- 44 words with more than one cluster
- 17 words with no meaningful sense distinction
- 6 words with faulty uses or clustering errors
- 20 words with named entities as the second cluster
  - 'Marco Polo'
- **1 word with a new sense**
  - 'broken heart' ('heart attack' vs metaphorical use)

Conclusion

- DURel pipeline is a viable tool to identify headwords that need to be updated
- given a large enough corpus, the method can be scaled to several thousand target headwords
- can be used to identify metaphorical/metonymic senses
- can be used to distinguish multiple named entities

Limitations

- small scale study
- PoS errors
- only tested on Germanic languages
- XL-Lexeme might have limitations for non-European languages
- relies on the availability of a written corpus

References I

Berlin-Brandenburgischen Akademie der Wissenschaften (Ed.). (2024). *DWDS - Digitales Wörterbuch der deutschen Sprache. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart.* Retrieved from https://www.dwds.de

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.

Fellbaum, C. (Ed.). (1998). *Wordnet: An electronic lexical database.*

Leipzig Corpora Collection. (2023a). *English news corpus based on material from 2023.* Leipzig Corpora Collection. Retrieved from https://corpora.uni-leipzig.de?corpusId=eng_news_2023_1M

Leipzig Corpora Collection. (2023b). *German news corpus based on material from 2023.* Leipzig Corpora Collection. Retrieved from https://corpora.uni-leipzig.de?corpusId=deu_news_2023_1M

OED. (2009). *Oxford English dictionary.* Oxford University Press.

Schlechtweg, D. (2023). *Human and computational measurement of lexical semantic change* (Doctoral dissertation, University of Stuttgart, Stuttgart, Germany). Retrieved from http://dx.doi.org/10.18419/opus-12833

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/

Sköldberg, E., Virk, S., Sander, P., Hengchen, S., & Schlechtweg, D. (2024). Revealing semantic variation in swedish using computational models of semantic proximity - results from lexicographical experiments. In *To appear.*