



University of Stuttgart
Germany



Testing the effect of using crowdsourced semantic proximity judgments in the process of human lexicographical word sense clustering

March 21, 2023

Lukas Theuer Linke

Bachelor's thesis

Institute for Natural Language Processing, University of Stuttgart

What does a lexicographer do?

Noun [\[edit \]](#)

project (*plural* **projects**)

1. A **planned endeavor**, usually with a specific **goal** and accomplished in several steps or stages. [\[quotations ▼\]](#)

2. (*usually in the plural, US*) An urban low-income housing building. [\[quotations ▼\]](#)

***Projects** like Pruitt-Igoe were considered irreparably dangerous and demolished.*

3. (*dated*) An idle scheme; an impracticable design.

*a man given to **projects***

4. (*US, sports*) a raw recruit who the team hopes will improve greatly with coaching; a **long shot diamond in the rough** [\[quotations ▼\]](#)

5. (*obsolete*) A **projectile**.

6. (*obsolete*) A **projection**.

Figure 1: A dictionary entry from Wiktionary¹

¹<https://en.wiktionary.org/wiki/project#Noun>

Motivation

- ▶ Using crowdsourced information in lexicographical tasks
- ▶ improving traditional methods
- ▶ Semantic proximity annotation: impact on lexicography

Related Work

- ▶ Text annotation tools
 - ▶ Different approaches
 - ▶ Lack of lexicographic capabilities
- ▶ Electronic lexicography
 - ▶ Tools
 - ▶ Potential of crowdsourcing

The Lexicographical Process in 4 steps

- 1.

The Lexicographical Process in 4 steps

1. collect uses of a word into a corpus,

The Lexicographical Process in 4 steps

1. collect uses of a word into a corpus,
2. cluster uses together based on their commonalities,

The Lexicographical Process in 4 steps

1. collect uses of a word into a corpus,
2. cluster uses together based on their commonalities,
3. discover what defines the commonality between the members of a cluster and

The Lexicographical Process in 4 steps

1. collect uses of a word into a corpus,
2. cluster uses together based on their commonalities,
3. discover what defines the commonality between the members of a cluster and
4. write a dictionary entry based on this definition.

Concordances

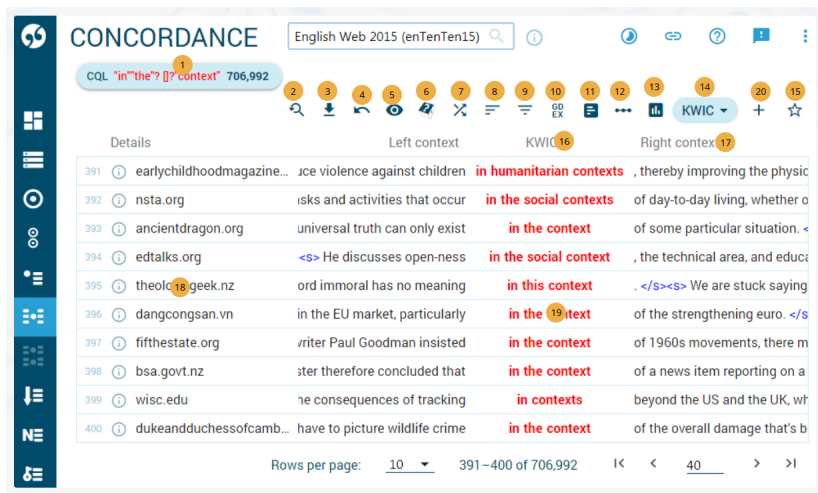


Figure 2: Concordance table for "context" from the online lexicography tool Sketch Engine (Kilgarriff et al., 2014)²

²<https://www.sketchengine.eu>

Semantic proximity annotation

- ▶ Based on the work of Schlechtweg (2023)
- ▶ Numerical scale for proximity
- ▶ Concept of WUGs - Word Usage Graphs

WUGs - example

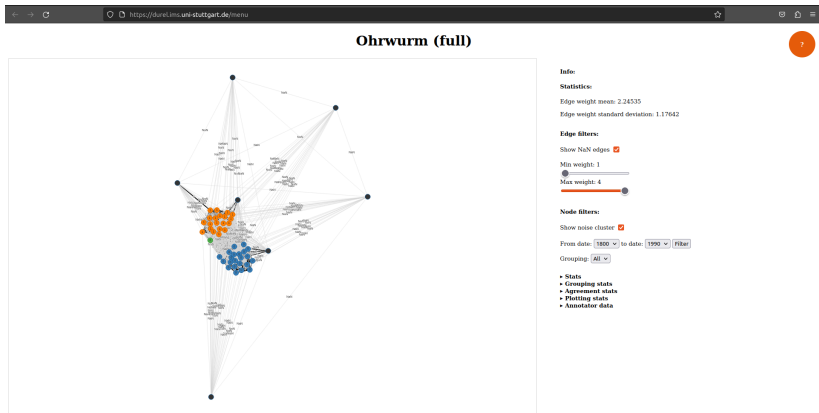


Figure 3: A WUG for "Ohrwurm"

WUGs

- ▶ Graphical representation of annotation
- ▶ Clusters of word meanings
- ▶ vertices represent word uses
- ▶ weights on edges represent the (median) semantic relatedness

DURel Tool

- ▶ Concordance tables
- ▶ Semantic proximity annotation
- ▶ Annotation statistics
- ▶ Visualization

Concordance tables

Left Context	Target	Right
Und zehn	Ohrwürmer	des deutschen Soul einspielte, englisch gesungen von
wei, so würde sich die gezähnte Hälfte eben wie der gespaltene	Ohrwurm	umkehren und den Hinterrest rein aufspeisen — und
Wenn Hansen den	Ohrwurm	von den Mädchen mit den roten Haaren singt, präsent
hrenqualle I. 132. 136. Ohrrobben II. 490. Ohrschnecken I. 356.	Ohrwurm	I. 532. Ohrwürmer I. 579. Oidemia II. 361. Oleniden I.
A Von dem	Ohrwurme	(Forficula); B von einer Heuschrecke (Locusta); C von
Im	Ohrwurm	ist der obere Magenmund inwendig wie mit einigen Z
Sie bilden eine Insektengruppe, die den	Ohrwürmern	und Heuschrecken näher steht als den Käfern.
	Ohrwurm	.
etou (7); 21.40 Sport am Abend; 22.45 Aktuelle Kamera; 23.00	Ohrwürmer	; 23.30 Nachrichten.
uns nähme ein Churschwerdt und schnitte ihn damit wie einen	Ohrwurm	entzwei, so würde sich die gezähnte Hälfte eben wie
Die Absicht der „rund“-Mannschaft, neben	Ohrwürmern	auch bislang weniger Bekanntes zu präsentieren, gin
Prominente Solisten singen einen „	Ohrwurm	“ nach dem anderen, das Fernseh-Ballett erinnert an
Oldies, doch immer noch	Ohrwürmer	
af einmal in anderen Insektenlarven, in denen von Blattwespen,	Ohrwürmern	, Käfern, vorherrschend jedoch in Schmetterlingsraup
Daß die	Ohrwürmer	ausschließlich Pflanzenfresser seien, glaube ich nich
	Ohrwurm	.
Männchen des großen	Ohrwurms	(Forficula gigantea).
Ein hartnäckiger	Ohrwurm	.
	Ohrwürmer	

Visualization

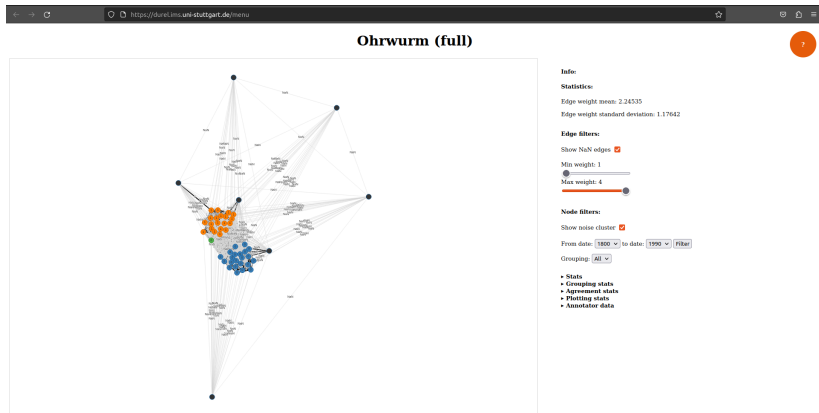


Figure 5: A WUG for "Ohrwurm"

User study

- ▶ Compares Use Analysis and Visualization
- ▶ Hypothesis: Visualization is better
 - ▶ Less common meanings are noticed
 - ▶ Consistent work
- ▶ Two study groups
- ▶ Additional questions on user satisfaction

Hypothesis

With the help of WUGs, humans can find more meanings of a word than using a traditional concordance table.

Study results

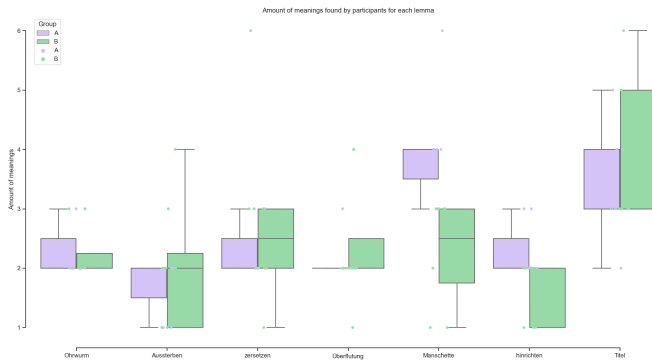


Figure 6: Meanings found by study participants

Conclusion

- ▶ WUGs improve results in lexicographical tasks somewhat
 - ▶ Concordances: some participants to overlook meanings
 - ▶ Informal feedback: tables are more tedious
- ▶ DUREl tool and WUGs need improvements

Future Work

- ▶ Optimization Work
- ▶ Additional features:
 - ▶ Copying texts from WUGs nodes
 - ▶ More intuitive overview of stats
- ▶ Studying the effect of WUGs with trained lexicographers

References I

- Kilgarriff, A., Baisa, V., Busta, J., Jakubíček, M., Kovár, V., Michelfeit, J., . . . Suchomel, V. (2014). The sketch engine: ten years on. *Lexicography*, 1, 7–36.
- Schlechtweg, D. (2023). *Human and computational measurement of lexical semantic change* (Unpublished doctoral dissertation). University of Stuttgart, Stuttgart, Germany.