



University of Stuttgart
Germany



Efficient Manual Word Sense Clustering on Historical Corpora

November 28, 2019

Dominik Schlechtweg

Supervisor: apl. Prof. Dr. Sabine Schulte im Walde

Institute for Natural Language Processing, University of Stuttgart, Germany

SemEval 2020 Task 1: Unsupervised Lexical Semantic Change Detection¹

- ▶ organized together with B. McGillivray, S. Hengchen, H. Dubossarsky, N. Tahmasebi
- ▶ evaluation is currently most pressing problem
- ▶ datasets for 4 languages
- ▶ comparison of two time periods t_1 and t_2

| | t_1 | t_2 |
|----------------|-------------------|-------------------|
| English | COHA 1810-1860 | COHA 1960-2010 |
| German | DTA 1800-1899 | BZ+ND 1946-1990 |
| Latin | LatinISE BC | LatinISE AD |
| Swedish | Kubhist 1790-1830 | Kubhist 1895-1903 |

Table 1: Time-defined subcorpora for each language.

¹<https://languagechange.org/semEval/>

SemEval 2020 Task 1: Unsupervised Lexical Semantic Change Detection

- ▶ two tasks:
 - ▶ **Task 1** – Binary classification: for a set of target words, decide which words lost or gained senses between t_1 and t_2 , and which ones did not.
 - ▶ **Task 2** – Ranking: rank a set of target words according to their degree of LSC between t_1 and t_2 .
- ▶ defined on **word sense frequency distributions**

Sense Frequency Distributions (SFDs)

| | t₁ | | | t₂ | | |
|---------------|----------------------|----------------|--------------|----------------------|----------------|--------------|
| Senses | Chamber | Biology | Phone | Chamber | Biology | Phone |
| # uses | 12 | 18 | 0 | 1 | 11 | 18 |

Figure 1: An example of a sense frequency distribution for the word *cell* in two time periods.

Annotation

- ▶ graded word sense annotation
- ▶ mostly based on **DURel** (Schlechtweg, Schulte im Walde, & Eckmann, 2018)

Scale


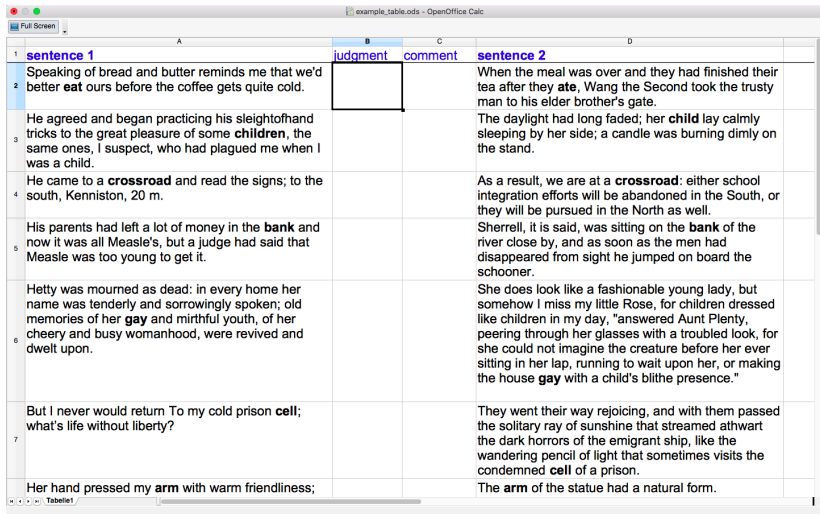
- 
- 4: Identical
 - 3: Closely Related
 - 2: Distantly Related
 - 1: Unrelated
- 0: Cannot decide

Table 2: Four-point scale of relatedness (Schlechtweg et al., 2018).

Data



The image shows a screenshot of an OpenOffice Calc spreadsheet titled "example_table.ods". The spreadsheet is in "Full Screen" mode and displays a table with 7 rows and 4 columns. The columns are labeled "A", "B", "C", and "D". The rows are numbered 1 through 7. The table content is as follows:

| | A | B | C | D |
|---|--|-----------------|----------------|---|
| 1 | sentence 1 | judgment | comment | sentence 2 |
| 2 | Speaking of bread and butter reminds me that we'd better eat ours before the coffee gets quite cold. | | | When the meal was over and they had finished their tea after they ate , Wang the Second took the trusty man to his elder brother's gate. |
| 3 | He agreed and began practicing his sleightofhand tricks to the great pleasure of some children , the same ones, I suspect, who had plagued me when I was a child. | | | The daylight had long faded; her child lay calmly sleeping by her side; a candle was burning dimly on the stand. |
| 4 | He came to a crossroad and read the signs; to the south, Kenniston, 20 m. | | | As a result, we are at a crossroad : either school integration efforts will be abandoned in the South, or they will be pursued in the North as well. |
| 5 | His parents had left a lot of money in the bank and now it was all Measle's, but a judge had said that Measle was too young to get it. | | | Sherrell, it is said, was sitting on the bank of the river close by, and as soon as the men had disappeared from sight he jumped on board the schooner. |
| 6 | Hetty was mourned as dead: in every home her name was tenderly and sorrowingly spoken; old memories of her gay and mirthful youth, of her cheery and busy womanhood, were revived and dwelt upon. | | | She does look like a fashionable young lady, but somehow I miss my little Rose, for children dressed like children in my day, "answered Aunt Plenty, peering through her glasses with a troubled look, for she could not imagine the creature before her ever sitting in her lap, running to wait upon her, or making the house gay with a child's blithe presence." |
| 7 | But I never would return To my cold prison cell ; what's life without liberty? | | | They went their way rejoicing, and with them passed the solitary ray of sunshine that streamed athwart the dark horrors of the emigrant ship, like the wandering pencil of light that sometimes visits the condemned cell of a prison. |
| | Her hand pressed my arm with warm friendliness; | | | The arm of the statue had a natural form. |

Table 3: Annotation Table.

Scale

| | A | B | C | D |
|---|---|----------|---------|---|
| 1 | sentence 1 | judgment | comment | sentence 2 |
| 2 | Speaking of bread and butter reminds me that we'd better eat ours before the coffee gets quite cold. | 4 | | When the meal was over and they had finished their tea after they ate , Wang the Second took the trusty man to his elder brother's gate. |

Figure 2: rating 4 (Identical).

Scale

| | A | B | C | D |
|---|--|-----------------|----------------|---|
| 1 | sentence 1 | judgment | comment | sentence 2 |
| 3 | He agreed and began practicing his sleightofhand tricks to the great pleasure of some children , the same ones, I suspect, who had plagued me when I was a child. | 3 | | The daylight had long faded; her child lay calmly sleeping by her side; a candle was burning dimly on the stand. |

Figure 3: rating 3 (Closely Related).

Scale

| | A | B | C | D |
|---|--|----------|---------|---|
| 1 | sentence 1 | judgment | comment | sentence 2 |
| 4 | He came to a crossroad and read the signs; to the south, Kenniston, 20 m. | 2 | | As a result, we are at a crossroad : either school integration efforts will be abandoned in the South, or they will be pursued in the North as well. |

Figure 4: rating 2 (Distantly Related).

Scale

| | A | B | C | D |
|---|---|-----------------|----------------|--|
| 1 | sentence 1 | judgment | comment | sentence 2 |
| 5 | His parents had left a lot of money in the bank and now it was all Measle's, but a judge had said that Measle was too young to get it. | 1 | | Sherrell, it is said, was sitting on the bank of the river close by, and as soon as the men had disappeared from sight he jumped on board the schooner. |

Figure 5: rating 1 (Unrelated).

Diachronic Data

- (1) 1830 but I am bound and thrown into a dark **cell**.
- (2) 1851 ...be fit to burn in a jail; no, not in a condemned **cell**.
- ...
- (3) 1990 But I never would return To my cold prison **cell**.
What's life without liberty?
- (4) 2006 She searched the bag for her **cell** as we headed toward the door.

Diachronic Data

| | A | B | C | D |
|---|---|-----------------|----------------|--|
| 1 | sentence 1 | judgment | comment | sentence 2 |
| 2 | but I am bound and thrown into a dark cell . | | | be fit to burn in a jail; no, not in a condemned cell . |
| 3 | but I am bound and thrown into a dark cell . | | | But I never would return To my cold prison cell ; what's life without liberty? |
| 4 | but I am bound and thrown into a dark cell . | | | She searched the bag for her cell as we headed toward the door. |
| 5 | be fit to burn in a jail; no, not in a condemned cell . | | | But I never would return To my cold prison cell ; what's life without liberty? |
| 6 | be fit to burn in a jail; no, not in a condemned cell . | | | She searched the bag for her cell as we headed toward the door. |
| 7 | She searched the bag for her cell as we headed toward the door. | | | But I never would return To my cold prison cell ; what's life without liberty? |

Table 4: Annotation Table.

Diachronic Data

| | A | B | C | D |
|---|---|-----------------|----------------|--|
| 1 | sentence 1 | judgment | comment | sentence 2 |
| 2 | but I am bound and thrown into a dark cell . | 4 | | be fit to burn in a jail; no, not in a condemned cell . |
| 3 | but I am bound and thrown into a dark cell . | 4 | | But I never would return To my cold prison cell ; what's life without liberty? |
| 4 | but I am bound and thrown into a dark cell . | 2 | | She searched the bag for her cell as we headed toward the door. |
| 5 | be fit to burn in a jail; no, not in a condemned cell . | 4 | | But I never would return To my cold prison cell ; what's life without liberty? |
| 6 | be fit to burn in a jail; no, not in a condemned cell . | 2 | | She searched the bag for her cell as we headed toward the door. |
| 7 | She searched the bag for her cell as we headed toward the door. | 2 | | But I never would return To my cold prison cell ; what's life without liberty? |

Table 5: Annotation Table.

From DUREl to SFDs

| | A | B | C | D |
|---|---|----------|---------|--|
| 1 | sentence 1 | judgment | comment | sentence 2 |
| 2 | but I am bound and thrown into a dark cell . | 4 | | be fit to burn in a jail; no, not in a condemned cell . |
| 3 | but I am bound and thrown into a dark cell . | 4 | | But I never would return To my cold prison cell ; what's life without liberty? |
| 4 | but I am bound and thrown into a dark cell . | 2 | | She searched the bag for her cell as we headed toward the door. |
| 5 | be fit to burn in a jail; no, not in a condemned cell . | 4 | | But I never would return To my cold prison cell ; what's life without liberty? |
| 6 | be fit to burn in a jail; no, not in a condemned cell . | 2 | | She searched the bag for her cell as we headed toward the door. |
| 7 | She searched the bag for her cell as we headed toward the door. | 2 | | But I never would return To my cold prison cell ; what's life without liberty? |

Table 6: Annotation Table.

Usage graphs

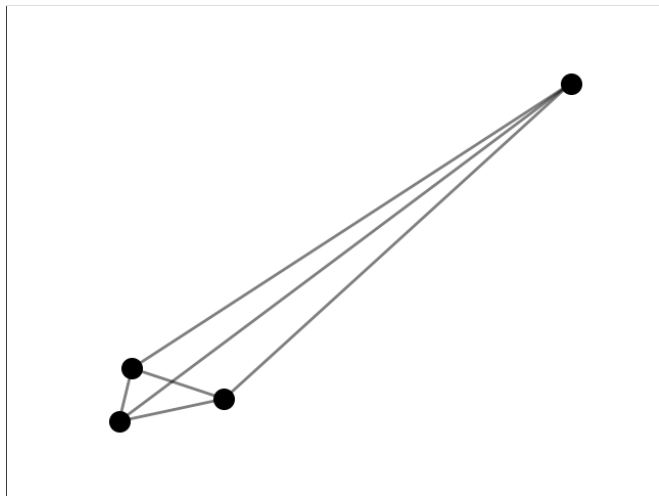


Figure 6: Graph visualization four uses of cell $T_1 = (2, 0)$ and $T_2 = (1, 1)$.

Time Information

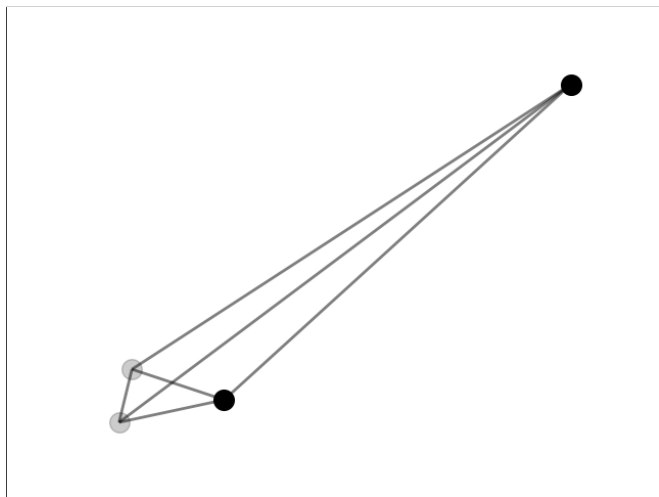


Figure 7: Graph visualization for uses of cell $T_1 = (2, 0)$ and $T_2 = (1, 1)$. Transparent nodes mark uses from t_1 .

Clustering

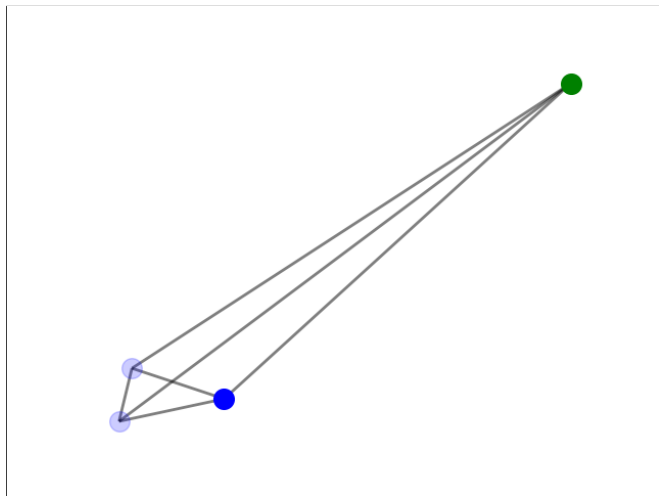


Figure 8: Graph visualization for uses of cell $T_1 = (2, 0)$ and $T_2 = (1, 1)$.

Advantages

- ▶ yields high inter-annotator agreement of non-experts
 - ▶ relies on intuitive linguistic concept of **semantic relatedness**
 - ▶ it is well-grounded in cognitive semantic theory
 - ▶ avoids assignment of particular sense to a word use
- requires only minimal preparation efforts

Imperfect Information

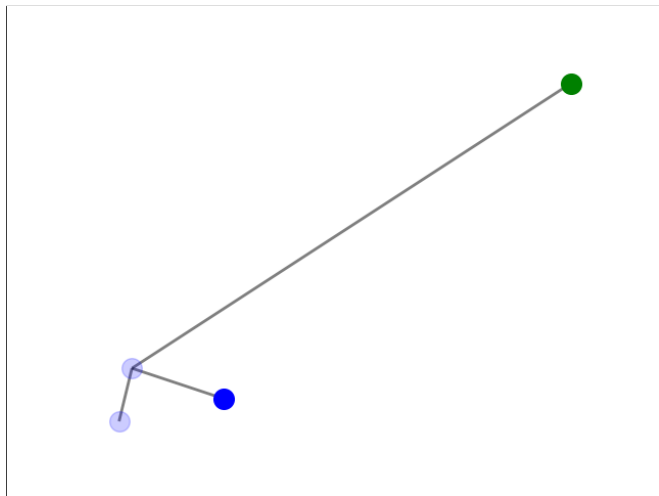


Figure 9: Graph visualization for uses of cell $T_1 = (2,0)$ and $T_2 = (1,1)$.

Main Disadvantages

- ▶ not feasible with huge edge number
- ▶ how to sample informative edges?

No information

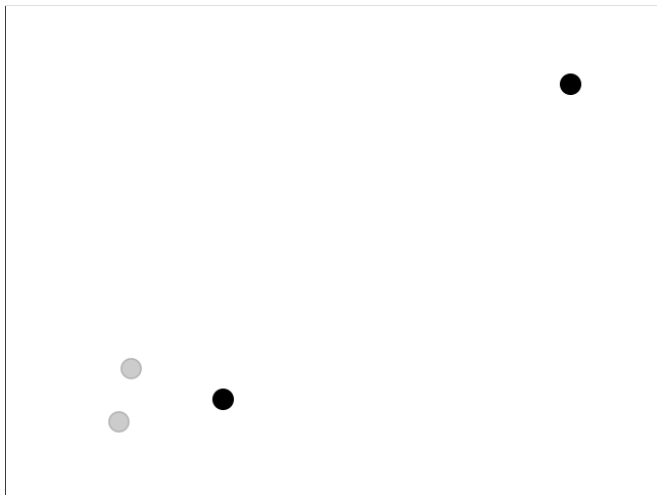


Figure 10: Graph visualization for uses of *cell*.

Edge sampling

- ▶ iteratively
- ▶ start with initial reference
- ▶ each subsequent round:
 1. get clustering
 2. combine uses with non-compared clusters
 3. combine non-compared clusters
- ▶ validated by simulation

Imperfect Information Graph

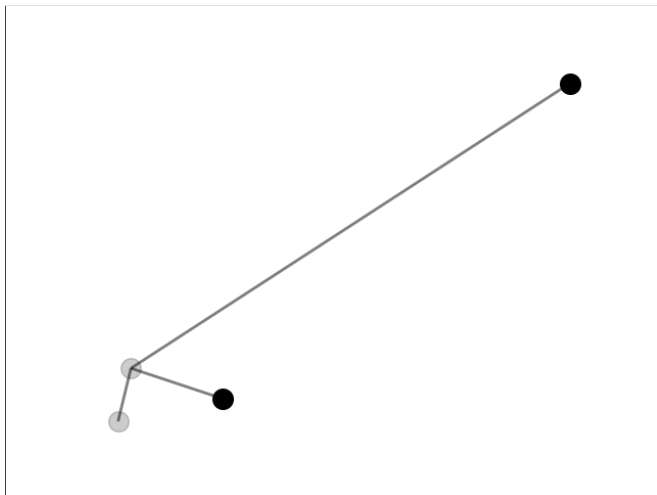


Figure 11: Graph visualization for uses of *cell*.

Clustering

- ▶ need for hard clustering
- ▶ (weighted) imperfect information graph
- ▶ optimization criterion: **reduce number of cluster-edge conflicts**
- ▶ simple approach:
 1. binarization with cutoff (two classes: (1,2) vs. (3,4))
 2. remove edges below cutoff (1,2)
 3. assume transitivity of relatedness relation
 4. clusters are connected components in reduced graph
 5. assume this is good solution
 6. optimization:
 - ▶ search for better solution by iteratively removing edges within clusters with conflicts

Clustering

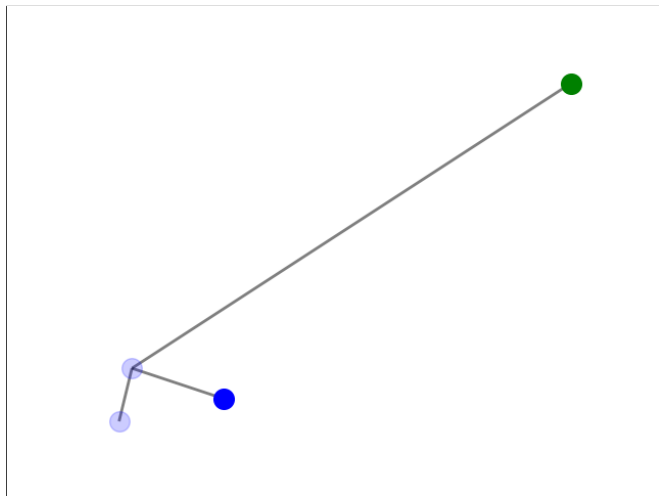


Figure 12: Graph visualization for uses of cell $T_1 = (2, 0)$ and $T_2 = (1, 1)$.

Output

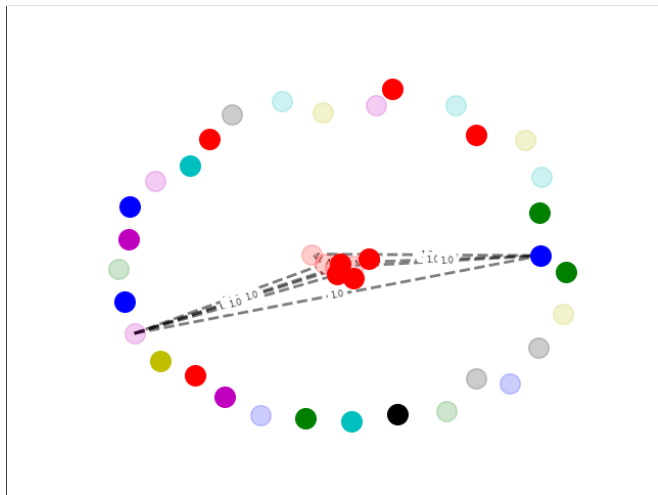


Figure 14: Step 1: Initial clustering

Output

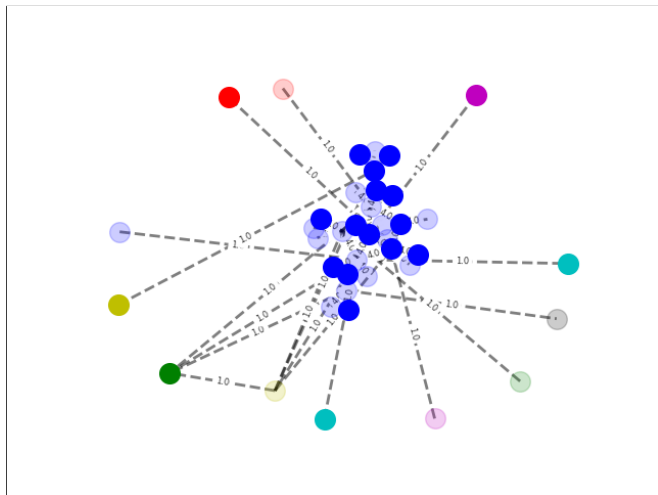


Figure 15: Step 2: Cluster comparison

Output

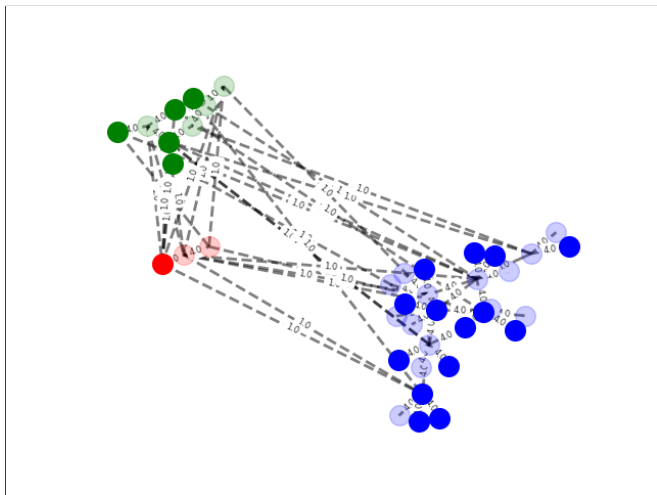


Figure 17: Step 4: Cluster comparison

Assumptions

- ▶ modern annotator can judge historical data
- ▶ good initial choice of reference
- ▶ clear-cut cluster structure
- ▶ transitivity

Bibliography

- 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* (p. 169-174). New Orleans, Louisiana.