Tracing and visualizing diachronic semantic change using contextualized embeddings

Software project, group 5

Averie (Ho Zoen) SO, NGO Van Duy, Scott TANKARD, Mathilde AGUIAR

Université de Lorraine

Monday 23 January 2023

Outline

Summary of work done

Updates

- full pipeline now working
- did re-training on larger corpus slices (SemEval EN, DE)
- progress on vizu component
- pipeline improvements
  - extracting pickles for semeval+gulordava wordlist
  - eval results (SemEval ranked graded.txt)
  - improved logging for g5_tools for unattended runs
- Decisions
  - alignment: concluded that it does not seem to be necessary
  - decided to not do google ngrams

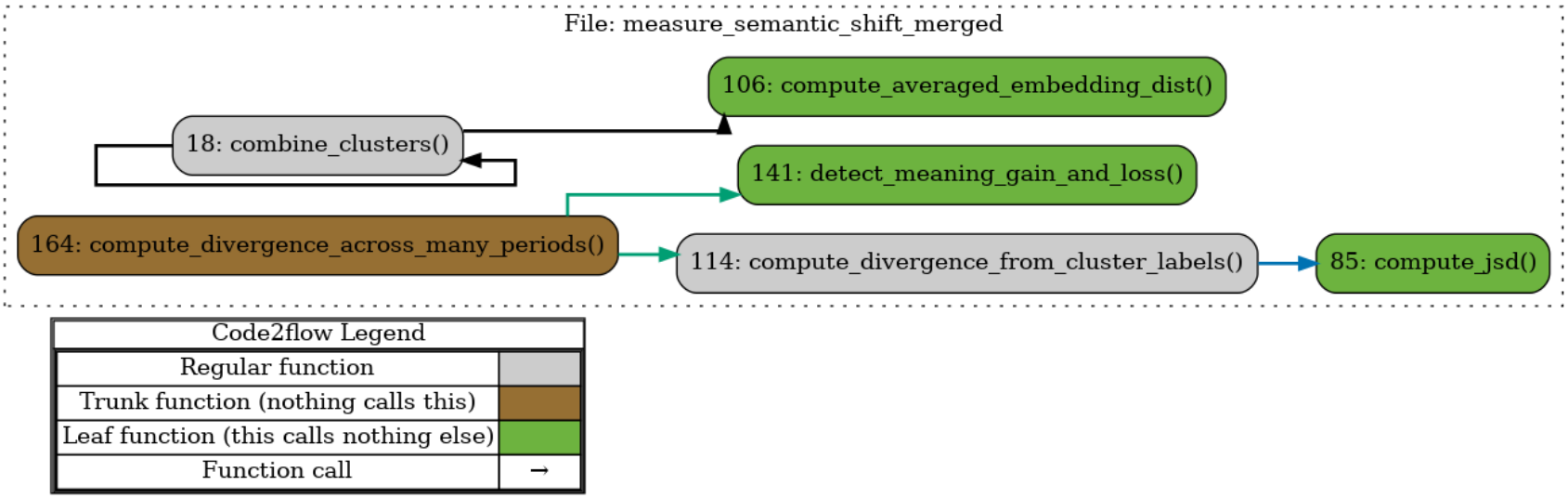


Figure: code2flow.measure_semantic_shift_merged.png

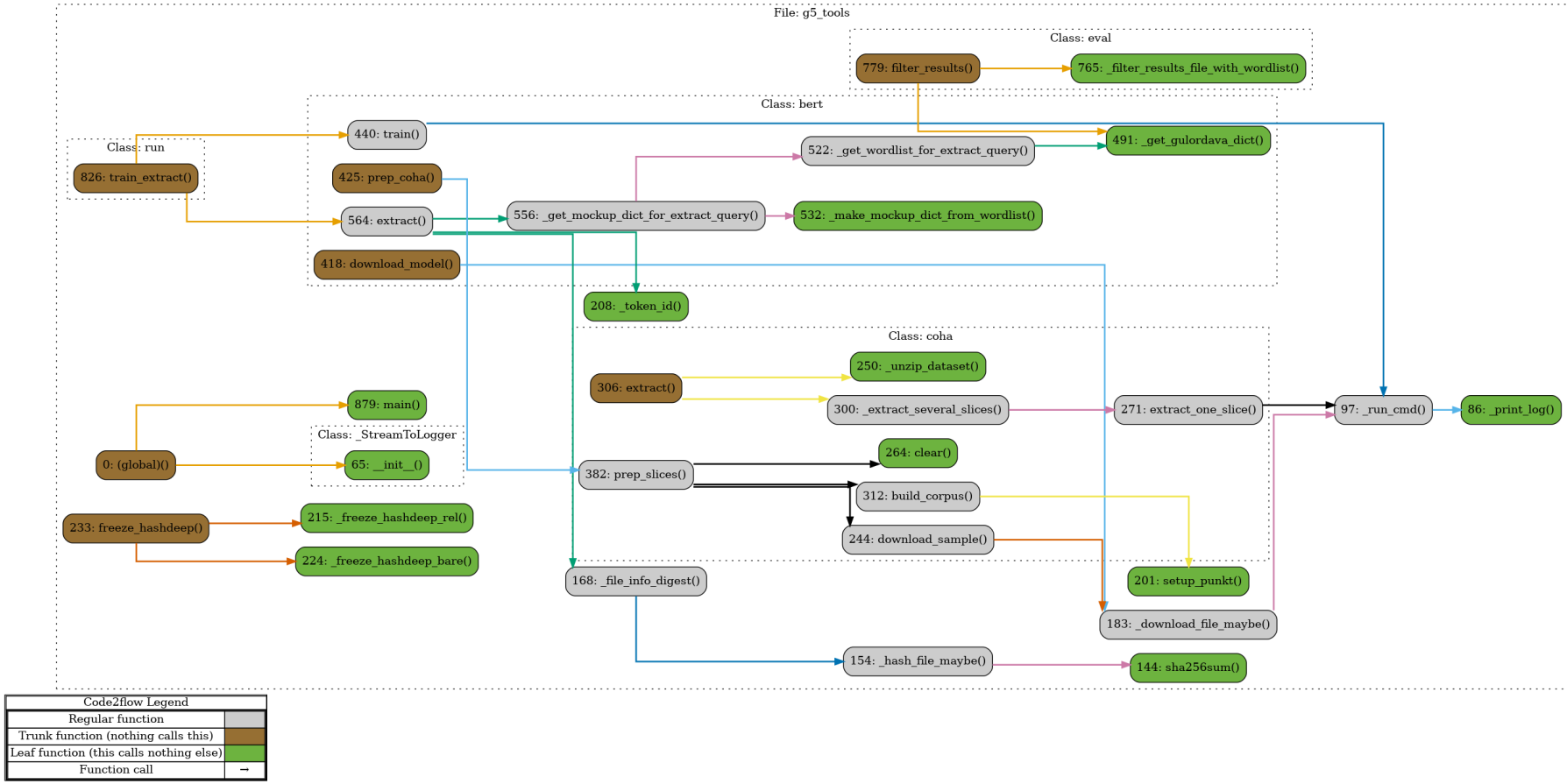


Figure: code2flow.g5_tools.png

Figure: code2flow.repo.png (scalable semantic shift repo)

Results

Evaluation on SemEval test set

attack_nn	0.1439699927	bag_nn	0.1003636619
ball_nn	0.4093665525	bit_nn	0.3065766263
chairman_nn	0	circle_vb	0.1710871199

Table: Examples of target words and the corresponding amount of semantic change.

model	**emb. type**	**English**	**German**
Schlechtweg et al (2019) [1]	SGNS	0.321	0.712
Pomsl and Lyapin (2020) [2]	SGNS	0.422	0.725
Montariol et al (2021) [3]	bert	0.456	0.583
our model (En only)	mbert	0.408	

Table: Spearman's **rank** correlation with human-annotated semantic change. 37 target words for English, 48 target words for German. Our models were trained on a reduced German dataset, 10 % of the original dataset.

[1] previous SOTA employing non-contextual word embeddings
[2] Winner of SemEval 2020 Task 1, subtask 2
[3] scalable_semantic_shift

Comparison with scalable_semantic_shift

model	k-means 5		k-means 7		AP	
	En	De	En	De	En	De
Montariol et al (2021)	0.375	0.520	-	-	0.437	0.561
our model (En only)	0.408		0.384		0.354	

Table: Spearman correlation with WD as measurement.

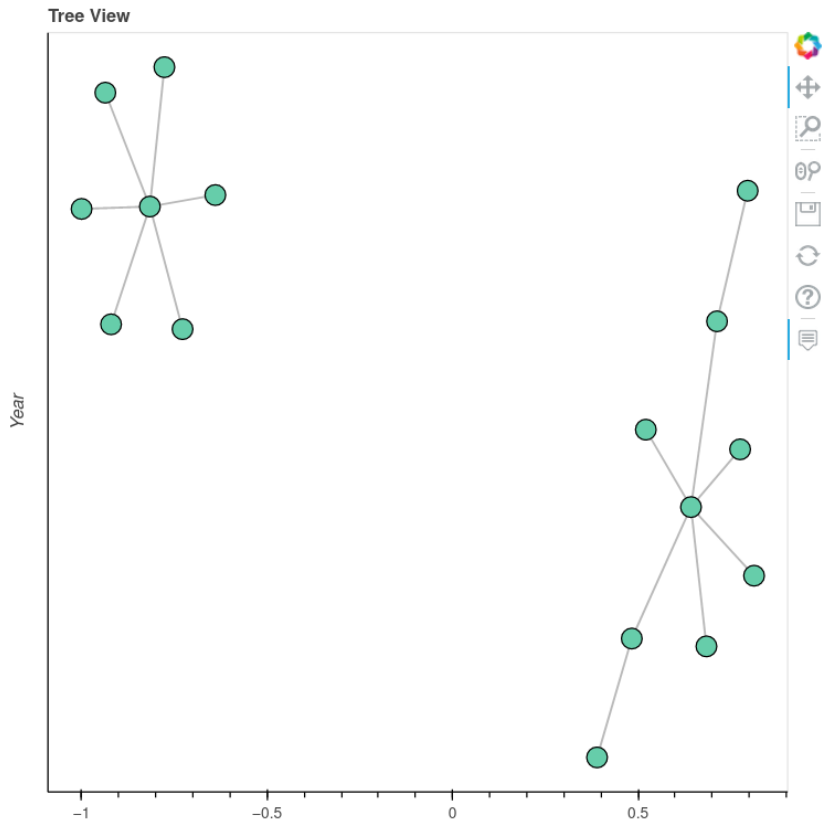
- Montariol et al (2021) experimented with two clustering methods (k-means, affinity propagation) and two change measurements (JSD, WD), and there doesn't seem to be one method that is clearly better across all languages
- Comparing with our results, the best performance for English is a k-means model with 5 clusters and using WD as measurement, which is not the same as the best English model in the original paper (affinity propagation + WD)

Visualization

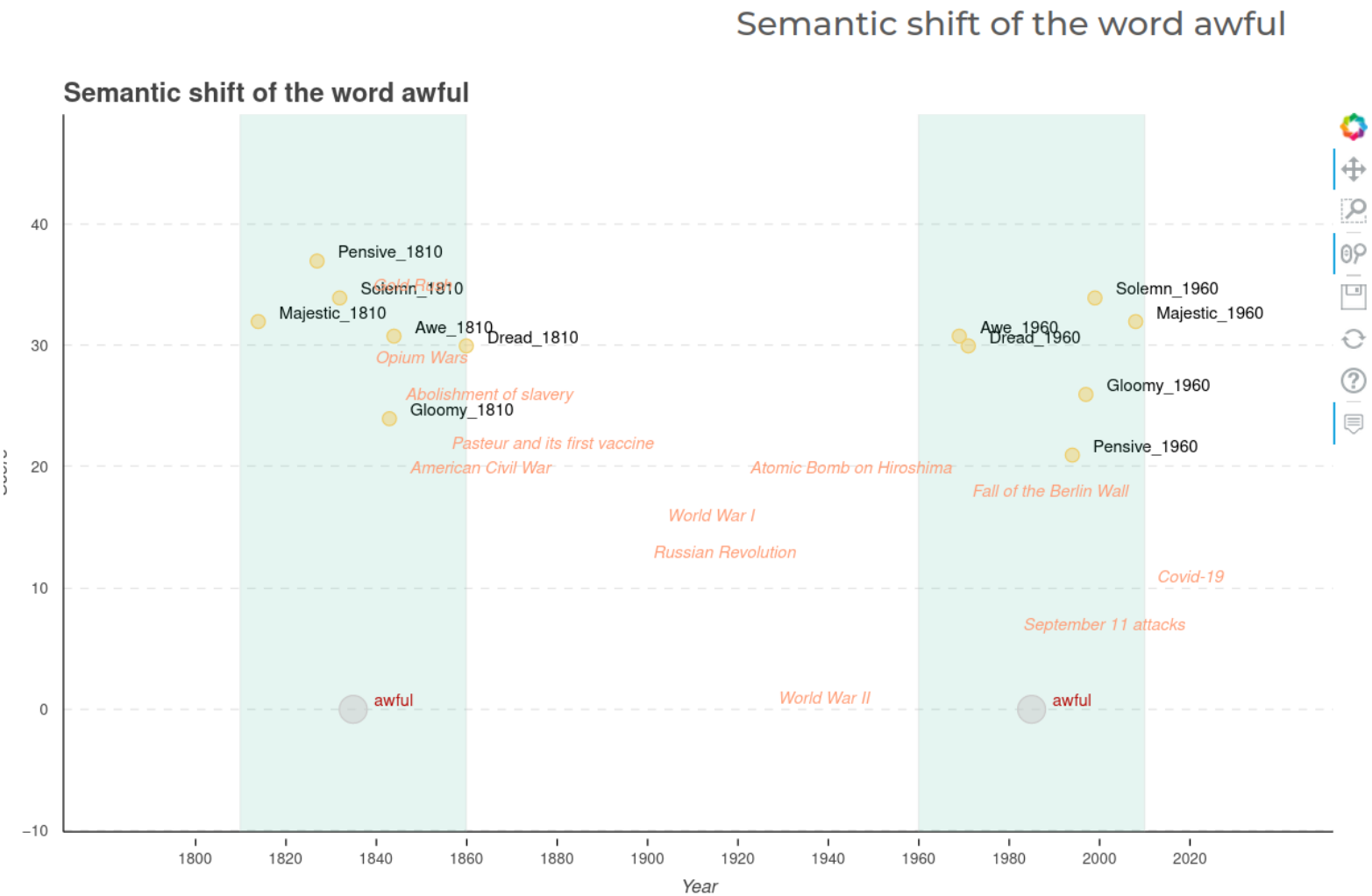
- New visualisation added: a tree visualization using the Python library *NetworkX*, still under construction
- Changes in the cluster visualization:
  - Historical events added
  - Visual representation of the time slices studied
- All the scripts to build the visualizations from the outputs CSV files
- Set of words available to query displayed on the interface

Visualization

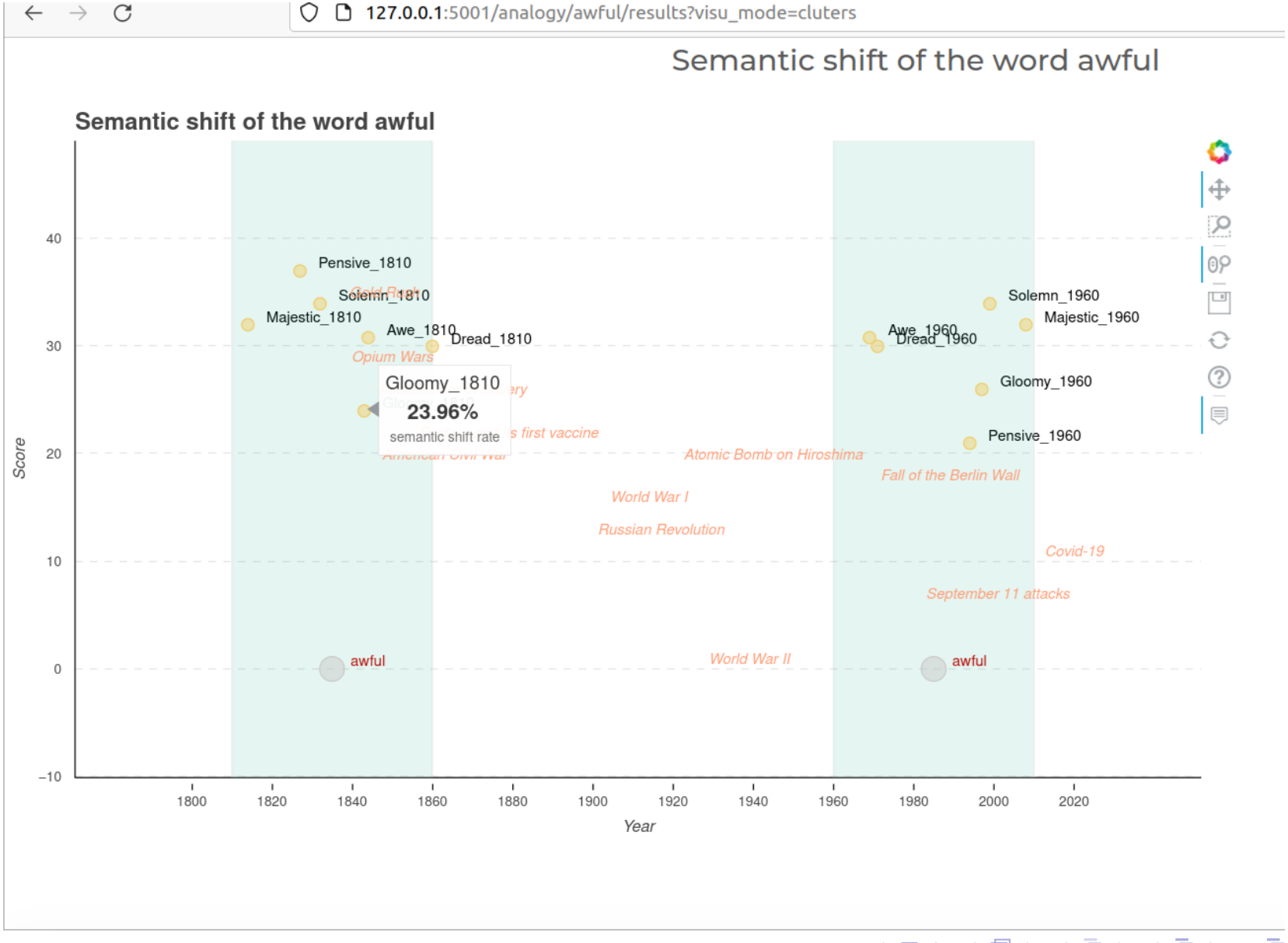
Semantic shift of the word awful



Visualization



Visualization



Ongoing work and challenges

Measuring Semantic Shift

Strategy:

- Export a limited set of words
- Get the k-nearest neighbours of the given words in the given slices
- Identify the meaning changes using the differences in neighbour sets

Works done:

- Get the `n` most frequent senses of the words
- Get the references of the senses

Works TBD:

- Measure the similarity of a word against the others

Web App

Remaining tasks:

- Change *measure_semantic_shift.py* to facilitate the generation of the tree visualization
- Implement the script to fetch the output CSV files from grid5k
- Deploy everything

Conclusion

Conclusions and next steps timeline

- Completed:
  - *understanding how to trace multiple senses in BERT; obtain corpora; further pre-train mBERT on two different periods of multilingual data; get program to generate quantified measurements of semantic change out of mBERT; prototype visualisation component.*
  - *re-train models on the larger corpus slices (from SemEval); evaluation with existing benchmarks (SemEval); implement alignment if necessary (decided not necessary); decide on practical trade-off solutions for visualisation UI and implement them; tweak training to improve eval results*
- TBD this week (before Thu 26 Jan): connect real results data into the visualisation UI; polish visualisation UI; writing report
- TBD Fri 27 Jan: turn in report

Thank you!

Question time