GeCKo, 18 May 2020
Integrating Generic and Contextual Knowledge

Analysing Lexical Semantic Change with Contextualised Word Representations

Mario Giulianelli, Marco Del Tredici, Raquel Fernández

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

University of Amsterdam

Types

Senses

highlighter

highlighter-pen

highlighter-makeup

?

?

Usages: contextualised representations

Number of usage types is **lexeme-specific** and **induced** from language use.

Usage vectors are **characterised by contexts of occurrence** — not by lists of nearest neighbouring words.

`<s> ... highlighter ... <\s>`

Method

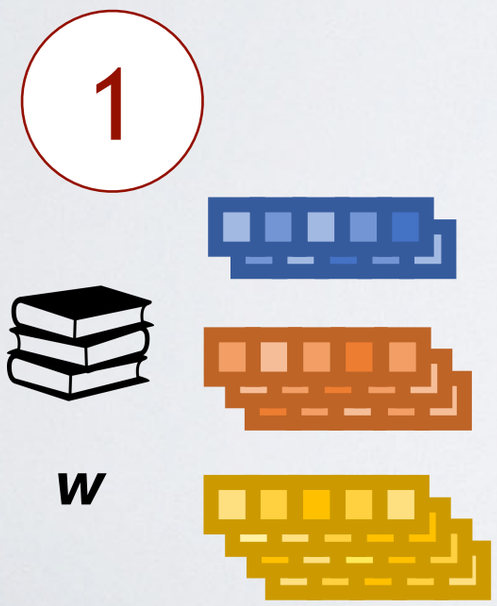
For each word of interest *w*

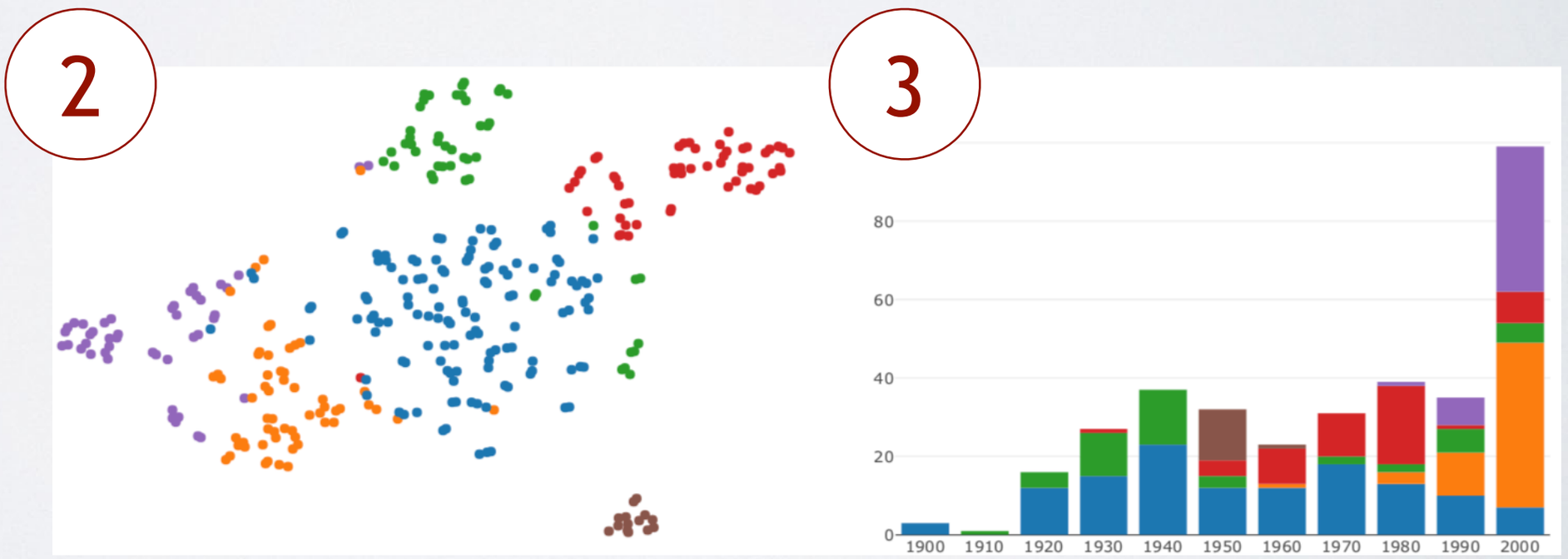
(1) **extract** contextualised representations for all occurrences of *w* in the corpus, using a language model (e.g., BERT or ELMo)

(2) **cluster** all representations of *w* into usage types by automatically selecting the optimal number of clusters (e.g. K-Means + silhouette score or Affinity Propagation)

(3) **organise** usage clusters into diachronic usage distributions (frequency-based or probability-based)

(4) **quantify** degree of change by comparing representations and usage distributions





Contextualised representations (left) and usage type distributions (right) for the word *users* as it occurs in COHA (Davies, 2012)

Method

For each word of interest *w*

(1) **extract** contextualised representations for all occurrences of *w* in the corpus, using a language model (e.g., BERT or ELMo)

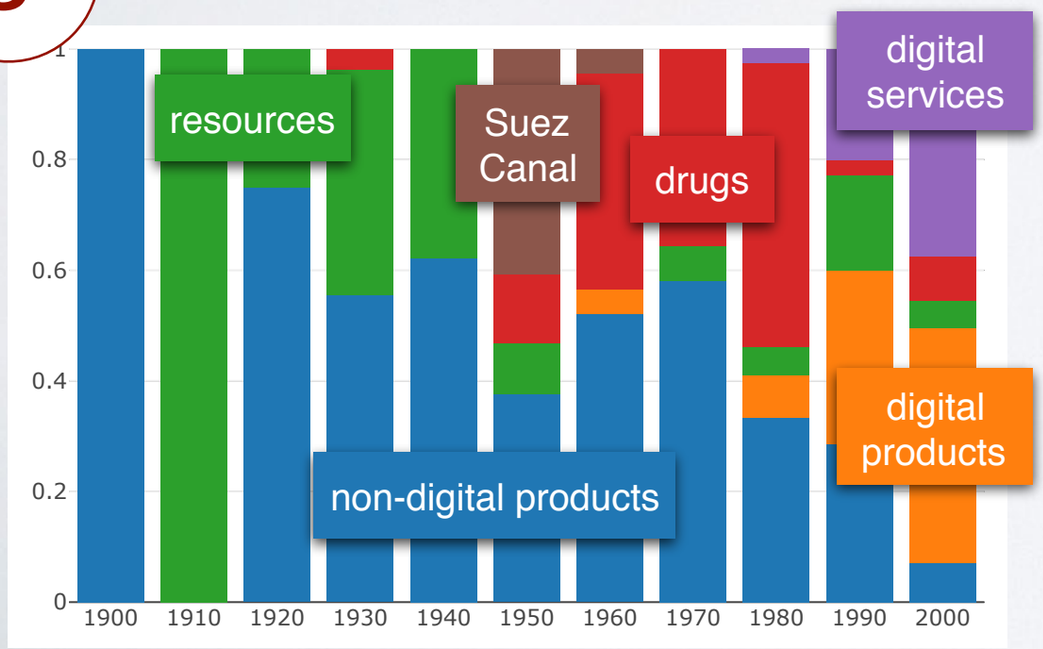
(2) **cluster** all representations of *w* into usage types by automatically selecting the optimal number of clusters (e.g. K-Means + silhouette score or Affinity Propagation)

(3) **organise** usage clusters into diachronic usage distributions (frequency-based or probability-based)

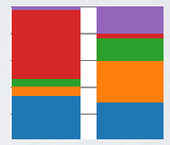
(4) **quantify** degree of change by comparing representations and usage distributions

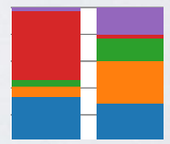
3

users of …



4

Jensen-Shannon Divergence ( *)*

Entropy Difference ( *)*

Average Pairwise Distance ( *)*

Are the resulting usage clusters interpretable?

*'the **ceiling** of a church'*

*'prefer the open sky to a **ceiling**'*

*'**ceiling** prices'*

*'breaking through the **ceiling**'*

literal vs metaphorical

polysemy and homonymy

*'full of questions, intensely **curious**'*

*'half fearful, half **curious**'*

*'the most **curious** reading'*

*'a **curious** sense of gratitude'*

*'**wireless** device'*

*'**wireless** network'*

entity names

*'**wireless**ly'*

*'verizon **wireless** theater'*

affixation

syntactic functionality

*'**refuse** to hire'*

*'**refuse** or neglect to perform'*

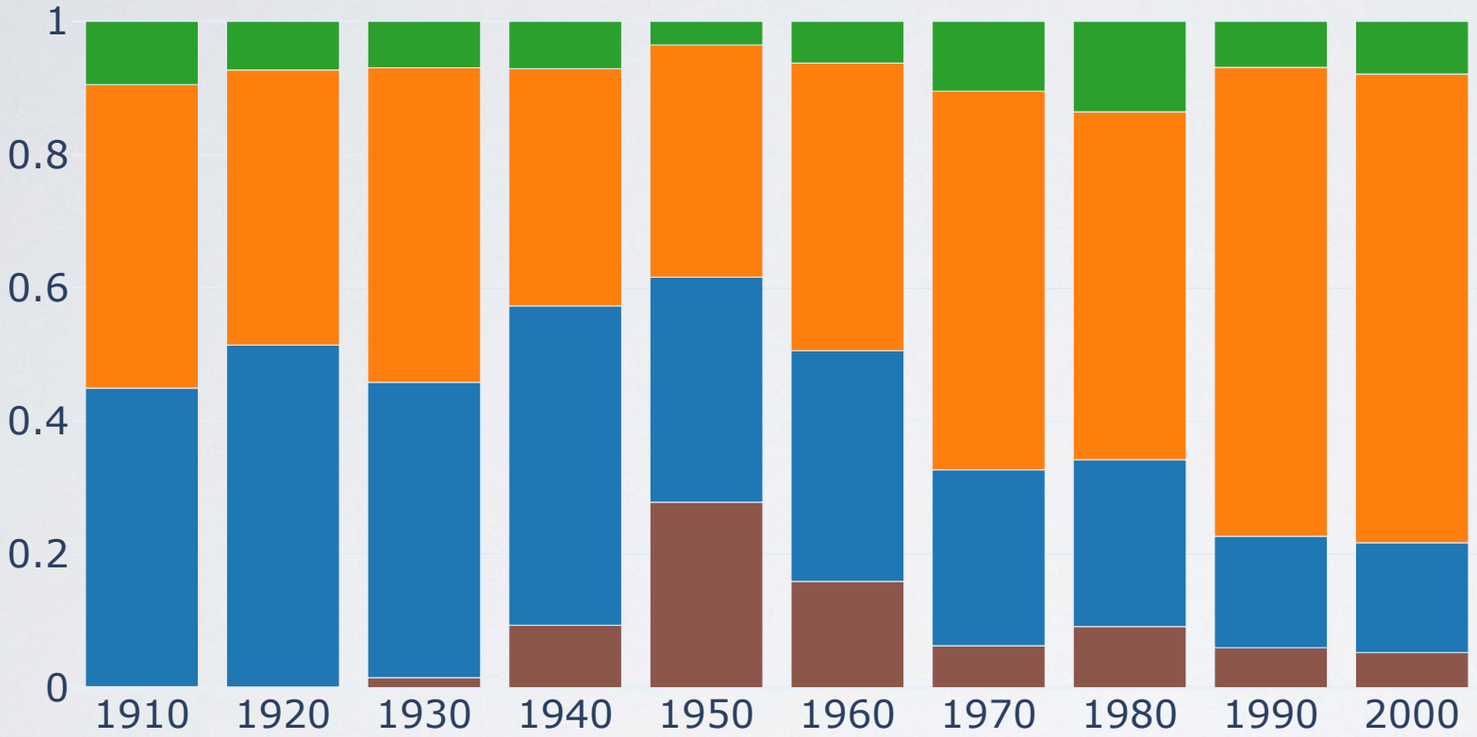
*'the **refuse** of the schools'*

*'**refuse** a draft'*

*'**refuse**, and you die'*

What types of lexical change are detected?

broadening (incl. metaphorisation)

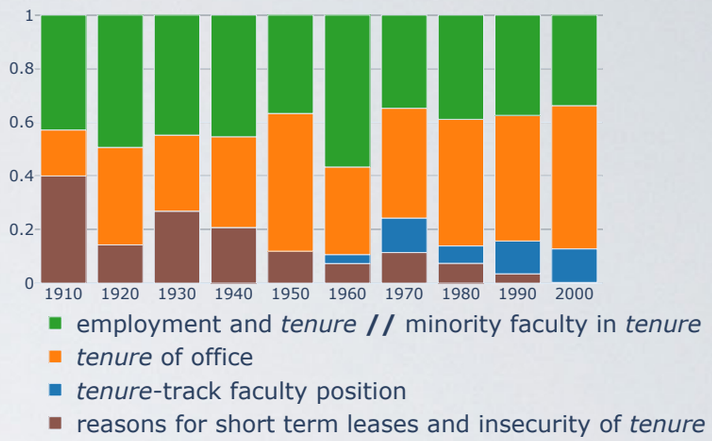


- I hung colored lights around my *curtain*less windows
- inflatable *curtain*-type head-protection bags
- raising the *curtain* on its [...] tax-reform program
- bureaucracies [...] on both sides of the *curtain*

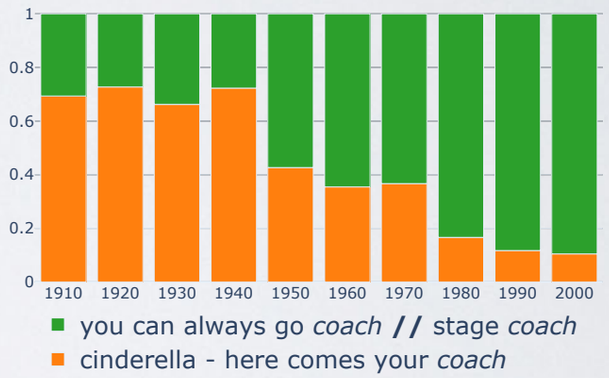
COHA (Davies, 2012)

COCA (Davies, 2010)

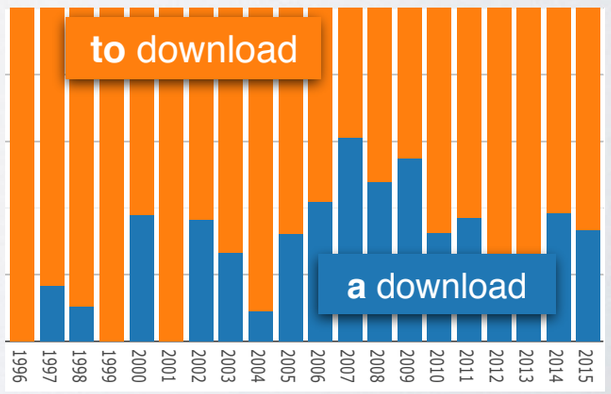
narrowing



shift



new syntactic role



Correlation with human judgements

Diachronic Usage Pair Similarity

A crowdsourced dataset of similarity judgements for more than 3K English word usage pairs (16 lemmas) from different time periods.

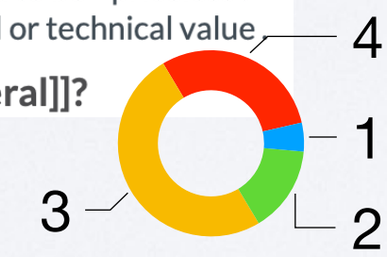
NEW DATASET: DUPS

federal

Please read carefully the following two sentences where the word [[federal]] occurs:

- robert m . hitchcock , who prosecuted the amerasia case in 1945 , testified today that he had been gravely handicapped because the government ' s best evidence had been produced by illegal seizures by [[federal]] agents . the prosecution , he asserted , was in fact fortunate under the circumstances to have done as well as it did .
- there should be such a fire every saturday afternoon at the same time with the same actual damage . this time it was the records and documents of the [[federal]] trade commisision , said to be " priceless . " also the reels of official motion pictures of historical or technical value

How similar are the two occurrences of [[federal]]?



Significant rank correlation between averaged human similarity judgements and BERT similarity scores for 10 out of 16 words.

Data: GEMS (Gulordava & Baroni, 2011) 100 words w/ shift scores.

Shift score: average human judgement on a word's meaning change between 1960 and 2000 (on a 4-points scale).

Metric: Spearman rank correlation between annotated change score and our three measures of change.

Frequency difference	0.068
Entropy difference (*max*)	0.278
Jensen-Shannon divergence (*max*)	0.276
Average pairwise distance (*Euclidean*, *max*)	0.285
Gulordava and Baroni (2011)	0.386
Frermann and Lapata (2016)	0.377

but wait for it...

Algorithm	*English*	*German*	*Latin*	*Swedish*
Word2vec CBOW cosine similarity baseline				
Incremental	0.210	0.145	0.217	-0.012
Procrustes	0.285	0.439*	0.387*	0.458*
Fine-tuned contextualised embeddings (top layer)				
ELMo Cosine similarity	0.254	**0.740***	0.360*	0.252
ELMo Average pairwise distance	**0.605***	0.560*	-0.113	**0.569***
BERT Cosine similarity	0.225	0.590*	**0.561***	0.185
BERT Average pairwise distance	0.546*	0.427*	0.372*	0.254

(Kutuzov and Giulianelli, 2020)

References

Davies, M. (2010). The 400-Million Word Corpus of Historical American English. *Corpora*.

Davies, M. (2012). The Corpus of Contemporary American English. *Literary* & *Linguistic Computing*.

Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL*.

Frermann, L., and Lapata, M. (2016). A Bayesian Model of Diachronic Meaning Change. *TACL*.

Gulordava, K., and Baroni, M. (2011). A Distributional Similarity Approach to the Detection of Semantic Change in the Google Books Ngram Corpus. In *Proceedings of the GEMS*.

Kutuzov, A., and Giulianelli, M. (2020). UiO-UvA at SemEval-2020 Task 1:Contextualised Embeddings for Lexical Semantic Change Detection. *Forthcoming*.