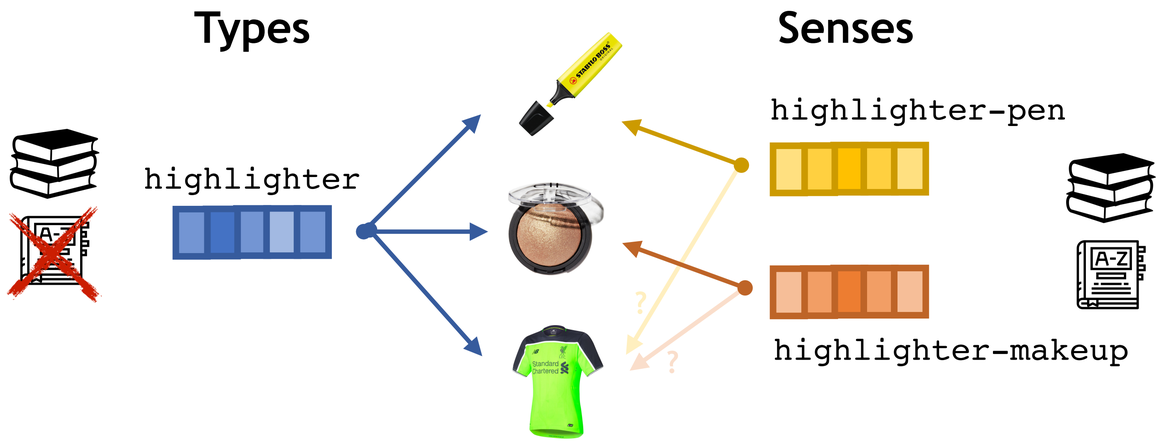
Lexical Semantic Change Analysis with Contextualised Word Representations

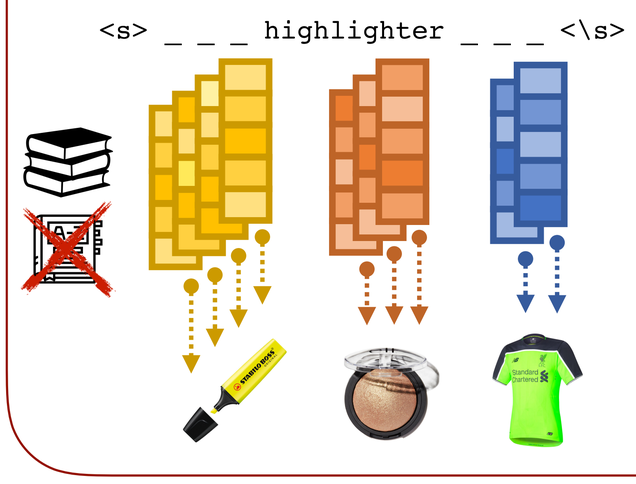
Mario Giulianelli, Marco Del Tredici, Raquel Fernández
Institute for Logic, Language and Computation, University of Amsterdam
{m.giulianelli|m.deltredici|raquel.fernandez}@uva.nl

By comparing clusters of **contextualised word representations**, we are able to detect **linguistic drifts** and **cultural shifts**.

Why contextualised representations?



Usages: contextualised representations



Number of usage types (≃ word senses) is **lexeme-specific** and **induced from language use**.

Usage vectors are **characterised by real sentences**—not by lists of nearest neighbouring words.

METHOD

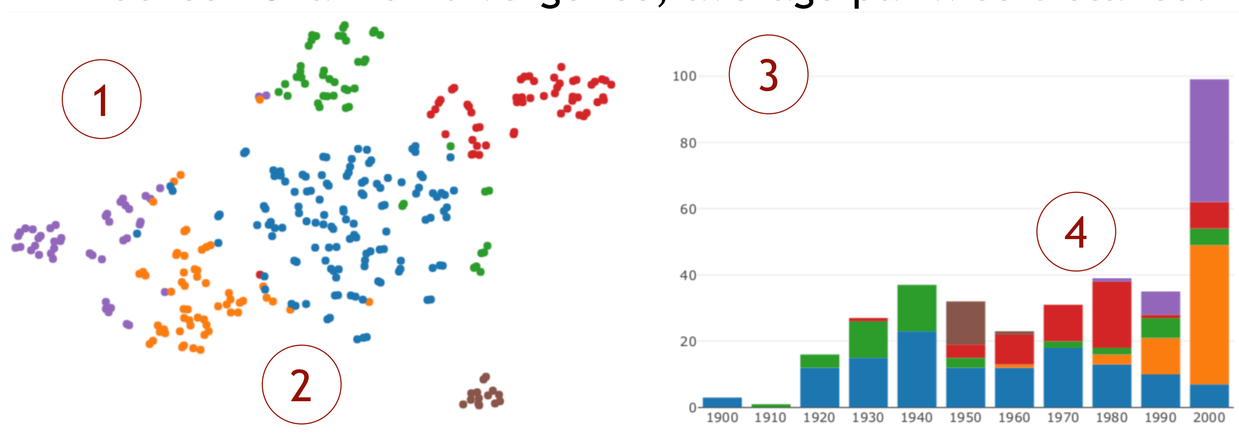
For each target word *w*

(1) **extract** contextualised representations for all occurrences of *w* in the corpus, using a language model (we rely on BERT; Devlin et al., 2019)

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

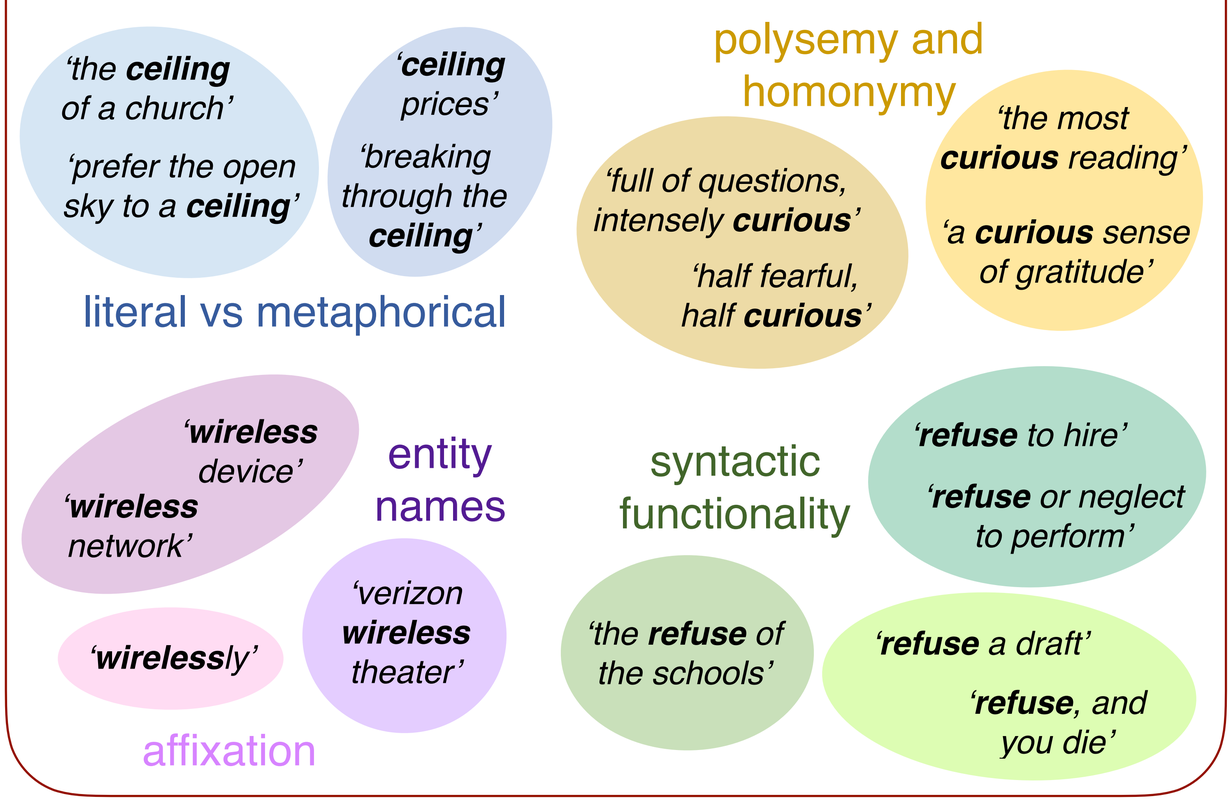
(3) **organise** clusters into diachronic usage distributions (frequency- or probability-based)

(4) **quantify** degree of change by comparing usage distributions, using three measures: entropy difference, Jensen-Shannon divergence, average pairwise distance.

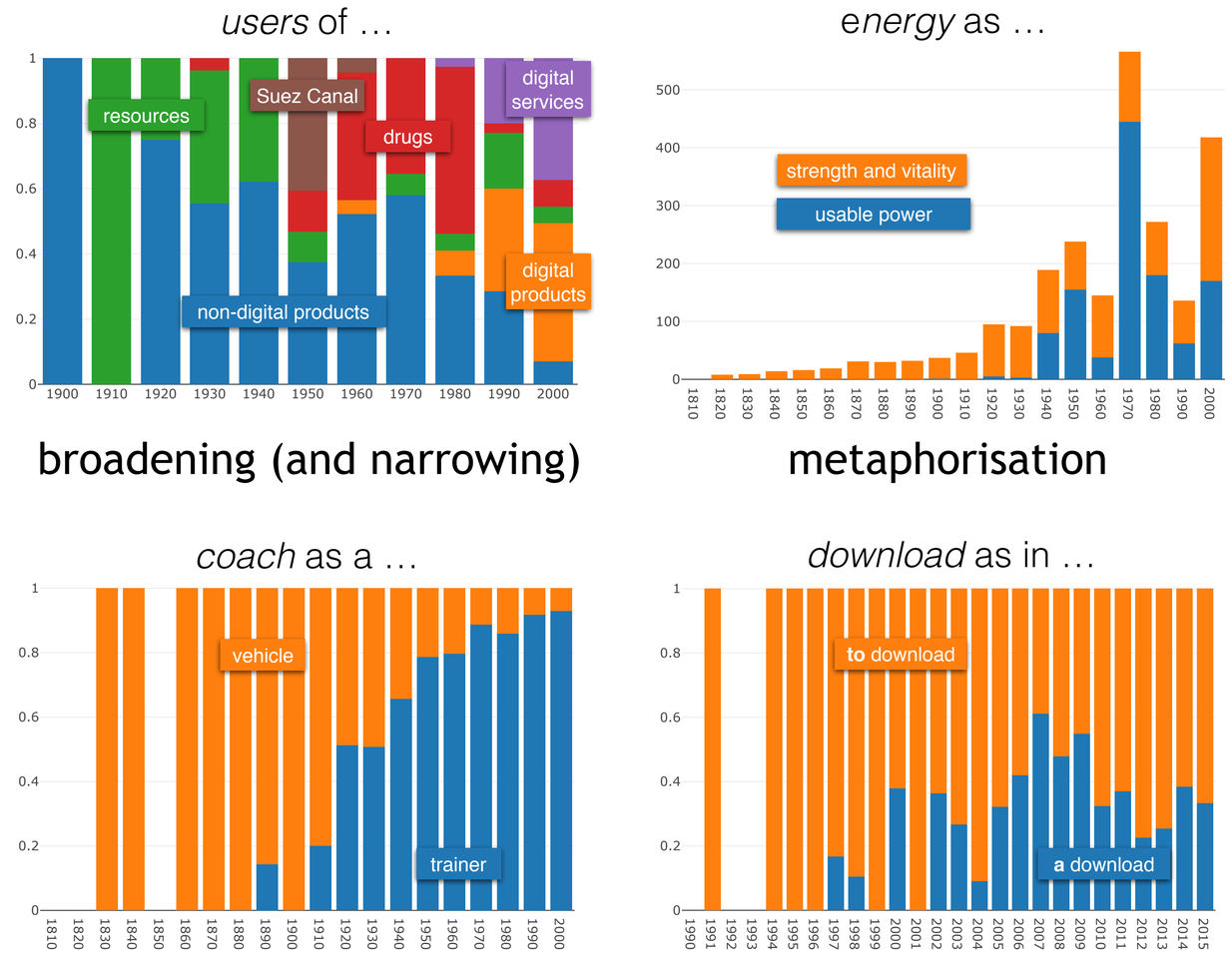


Contextualised representations (left) and usage type distributions (right) for the word *users* as it occurs in COHA (Davies, 2012)

Are the resulting usage clusters interpretable?



What types of lexical change are detected?



EVALUATION: Quantified vs perceived change

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

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

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

	Corpus	r	ρ
Gulordava & Baroni (2011)	Google Books	0.386	n.a.
Frermann & Lapata (2016)	DATE	n.a.	0.377
Skip-gram distance	COHA	0.047	0.119
Entropy difference	COHA	0.217	0.264
Mean distance	COHA	0.224	0.293
Jensen Shannon distance	COHA	0.231	0.224

DATA

COHA (Davies, 2012)
1810's, 1820's, …, 2000's

COCA (Davies, 2010)
1990, 1991, …, 2015

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., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL*.
Frermann, L., & Lapata, M. (2016). A Bayesian Model of Diachronic Meaning Change. *TACL*.
Gulordava, K., & Baroni, M. (2011). A Distributional Similarity Approach to the Detection of Semantic Change in the Google Books Ngram Corpus. In *Proceedings of the GEMS*.