

Modeling the evolution of English noun compounds with feature-rich diachronic compositionality prediction

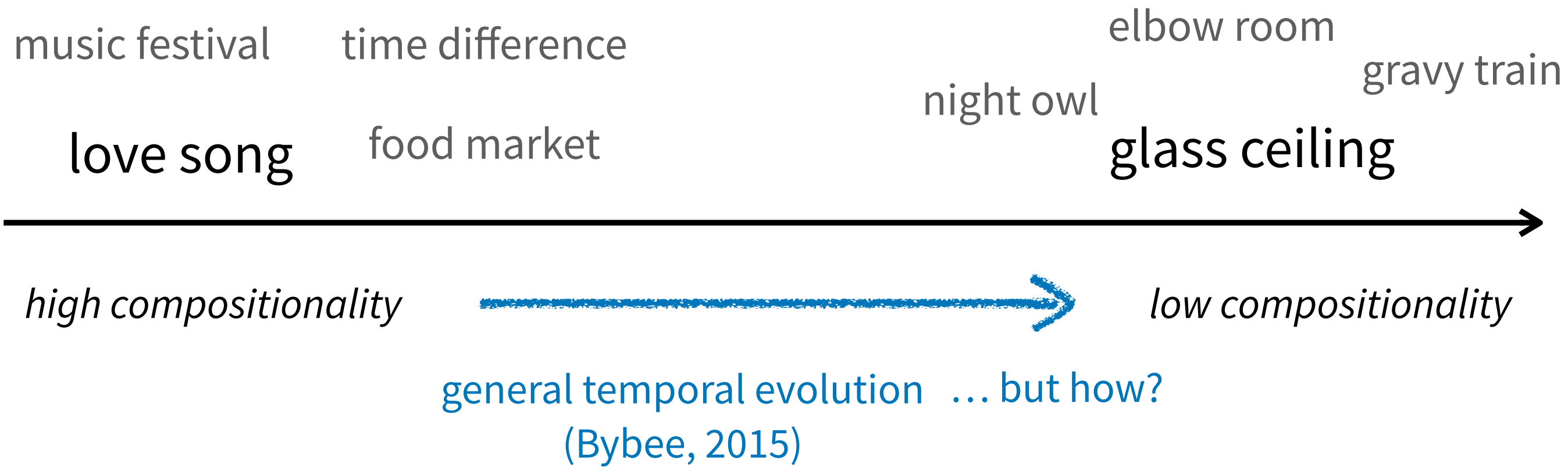
Filip Miletić & Sabine Schulte im Walde
 {filip.miletic, schulte}@ims.uni-stuttgart.de

ACL 2025
VIENNA

University of Stuttgart
 Institute for Natural Language
 Processing



Why do some noun compounds become (non-)compositional over time?

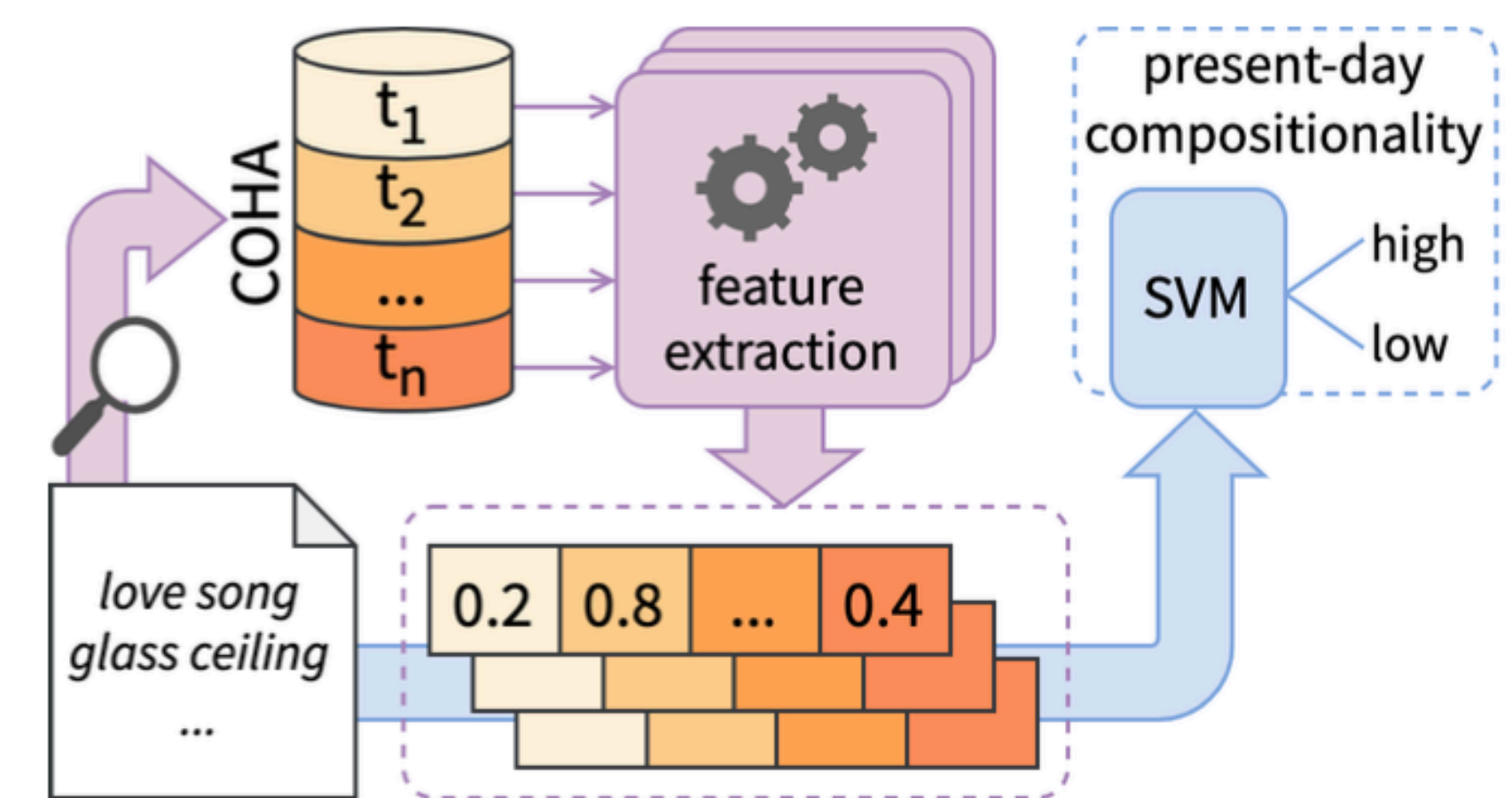


We predict **present-day compositionality** of English noun compounds using a diverse and linguistically motivated set of **time-specific features** reflecting the **evolution** of compound usage over time.

Experimental setup

Corpus: CCOHA → 1830-2009 (>400m words)
 (Davies, 2012; Alatrash et al., 2020)

Gold standard: 210 noun-noun compounds
 (Reddy et al., 2011; Cordeiro et al., 2019)



RQ1: Which diachronic properties are predictive of present-day compositionality?

Family	Approach	Accuracy		
		Comp	Modif	Head
Dispersion	frequency	0.631	0.608	0.619
	productivity	0.622	0.629	0.566
Static representations	cooccurrences	0.767	0.743	0.763
	word2vec	0.844	0.871	0.776
	topic model	0.746	0.793	0.693
Transformers (general)	BERT	0.849	0.748	0.792
	RoBERTa	0.726	0.720	0.724
	ALBERT	0.847	0.774	0.760
	DistilBERT	0.793	0.742	0.769
	SpanBERT	0.740	0.688	0.751
Transformers (historical)	histLM	0.716	0.704	0.688
	hmBERT	0.613	0.650	0.642
	MacBERTh	0.682	0.692	0.660
	MacBERTh-WSD	0.668	0.692	0.660
Random		0.500	0.500	0.500

Direct features

- Frequency
- Productivity

glass bottle *debt ceiling*
glass door *room ceiling*
glass window *plaster ceiling*

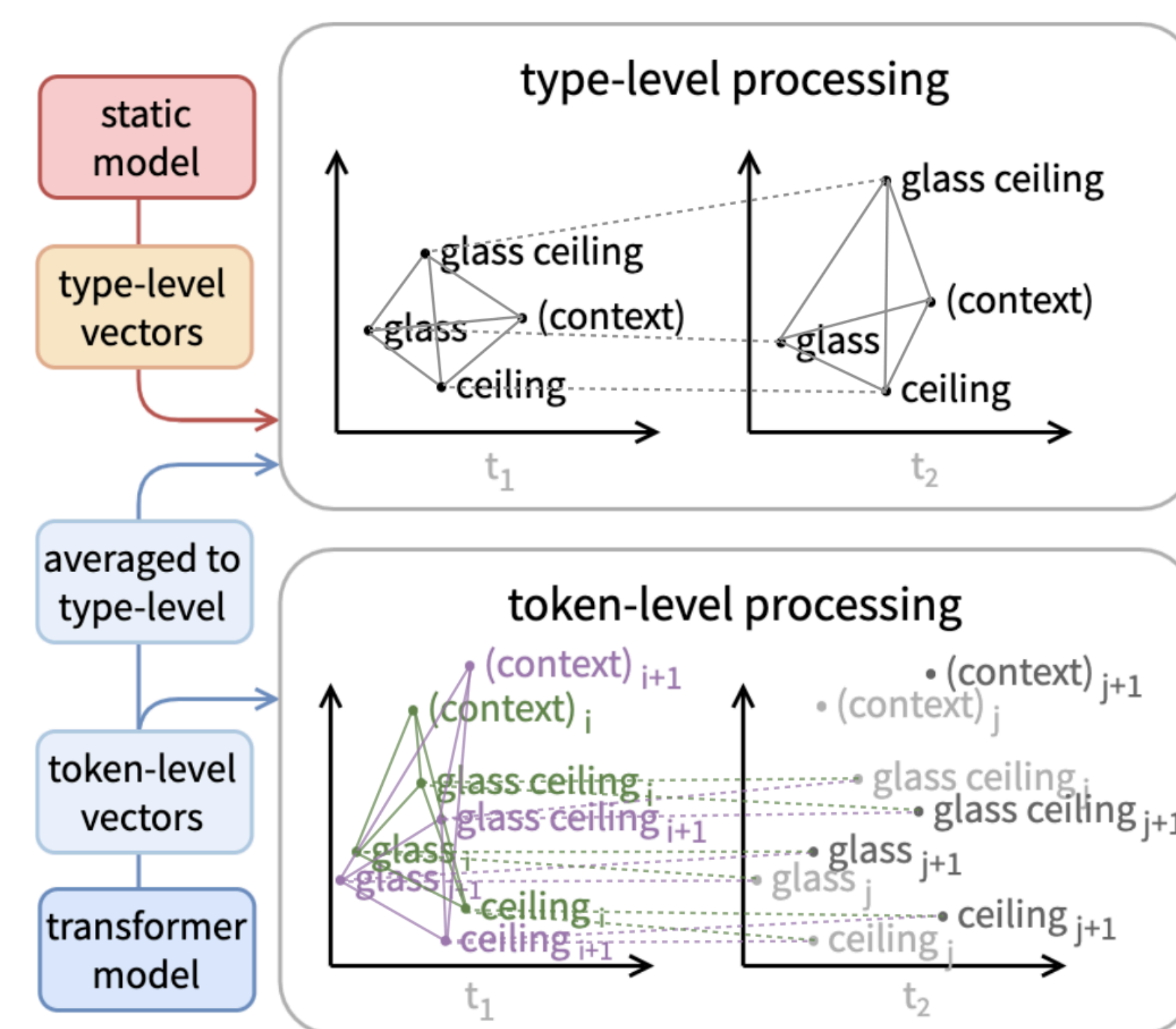
Derived features

Static models

- cooccurrences
- word2vec
- SBM

Transformer models

- BERT, RoBERTa
- DistilBERT, ALBERT
- SpanBERT
- hmBERT, histLM, MacBERTh, MacBERThWSD



Processing strategy
 → type-level
 → token-level

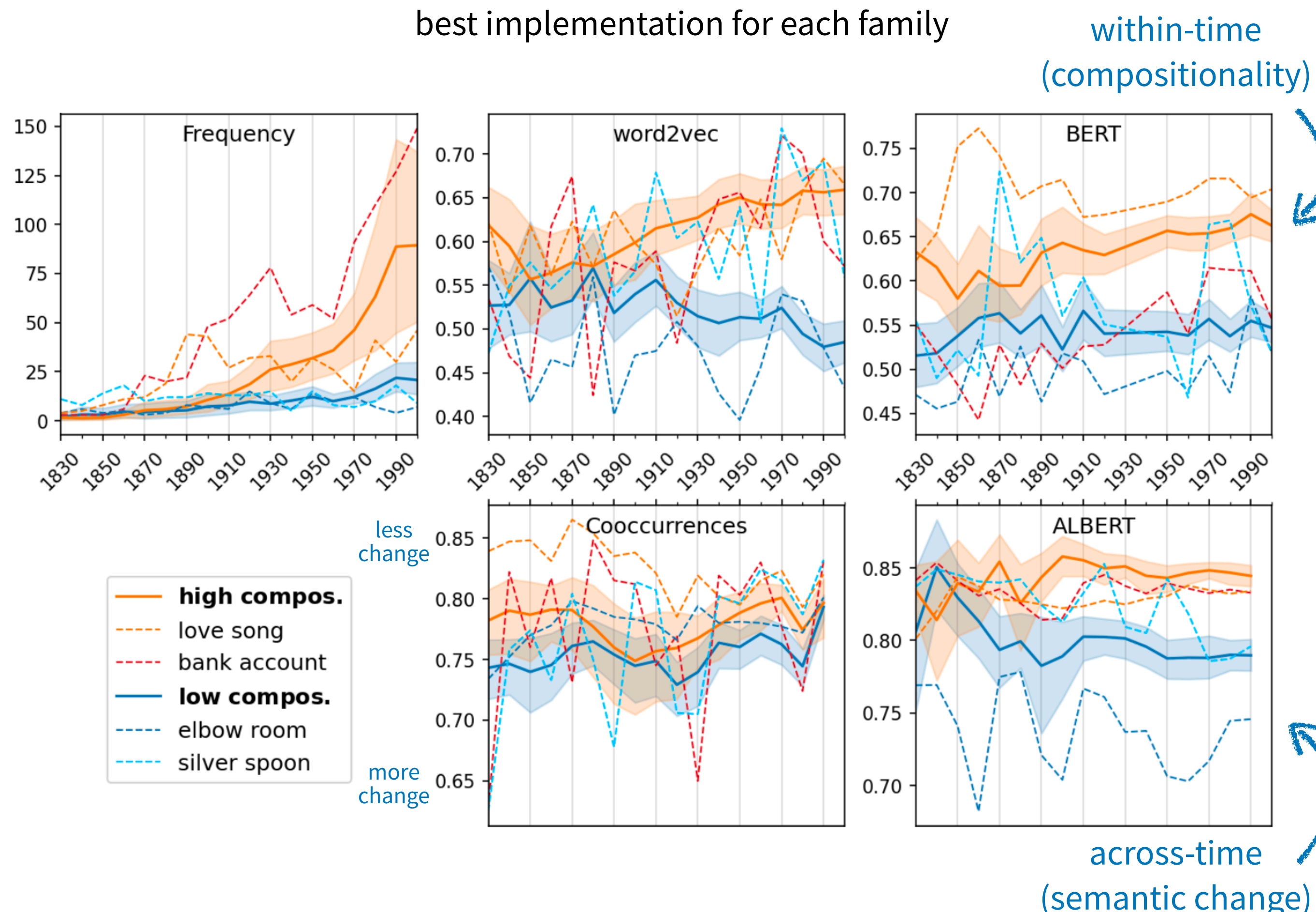
Target vectors
 → compound
 → modifier
 → head
 → context

Temporal settings
 → within-time: compositionality prediction
 → across-time: semantic change prediction

Similarity functions
 → cosine score
 → distributional nearest neighbors

RQ2: What is the relationship between compositionality and semantic change?

Evolution of feature values over time
 best implementation for each family



RQ3: How robust are different features to changes in modeling strategies?

Differences in performance based on transformer layers & estimate types

