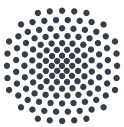
Classifying Noun Compounds for Present-Day Compositionality:

Contributions of Diachronic Frequency and Productivity Patterns

Maximilian Maurer, Chris Jenkins, Filip Miletić,
Sabine Schulte im Walde

September 19, 2023



University of Stuttgart
Institute for Natural Language Processing

Introduction: Compositionality

climate change

High compositionality

- Literal

melting pot

Low compositionality

- Nonliteral

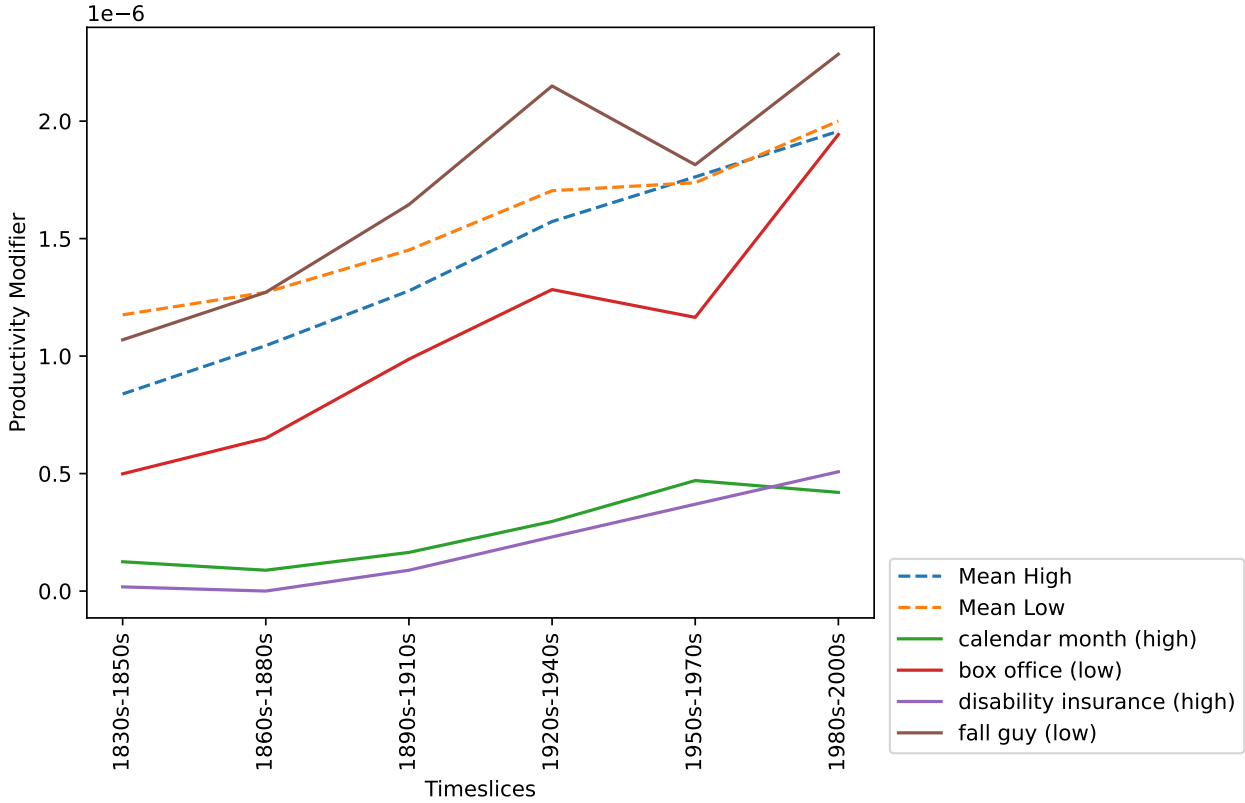
Motivation: Diachronic perspective

Hypothesis:
Differing patterns for (present-day) low- vs. high-compositionality compounds over time for

- high vs. low frequency

  [Lee (1990), Hamilton et al. (2016)]
- high vs. low productivity

  [Jurafsky et al. (2001), Hilpert (2015)]



Approach

- Assessing informativeness of patterns via binary classification
- Using empirical properties over time as features

$$[F_{1830s}, F_{1840s}, \ldots, F_{2000s}]$$

Experimental Setup

Data

- 185 human-annotated noun-noun compounds [Cordeiro et al. (2019)]
- Diachronic corpus: Clean Corpus of historical American English (CCOHA) [Alatrash et al. (2020)]
  - Data from 1830s-2000s from different text genres
- Synchronic corpus: ENCOW [Schäfer (2015)]

Empirical Properties

- Compound frequency
- Constituent frequency
- Productivity/Morphological family size

collected for decades and 30-year timeslices,
normalized per M tokens

Experimental Setup

Targets

- 62 most and 62 least compositional noun compounds (Leaving out 63 mid-scale compounds)
- For Compound, Modifier and Head compositionality
  $\rightarrow$ Three experiments

Classifier

- SVM
- Properties and combinations of properties as features

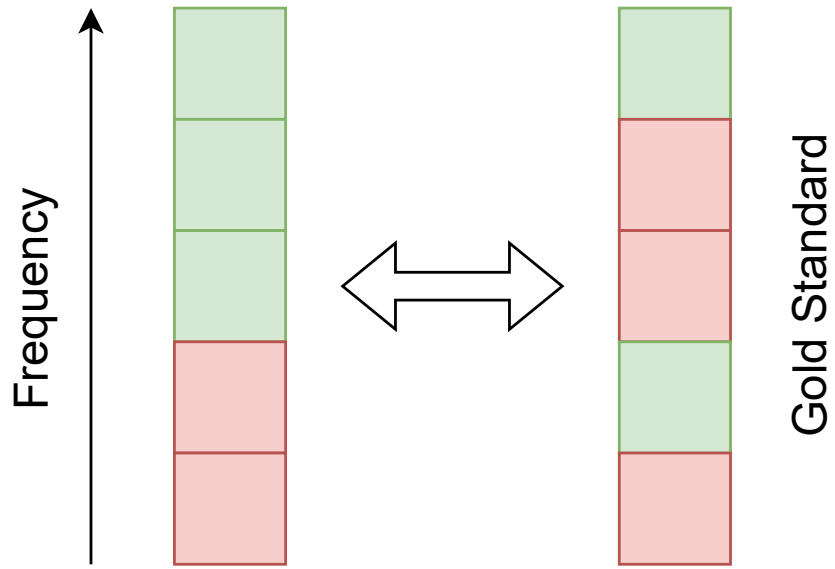
Evaluation

- Accuracy
- Repeated k-fold cross-validation (4 folds, 8 repetitions)

Static synchronic experiment

For ENCOW/last CCOHA timeslice

- Order target by property
- Assume top N compounds to be the highly compositional class
- Evaluate against gold standard compositionality classes



Results

Features	Accuracy					
	Compound		Modifier		Head	
	coarse	fine	coarse	fine	coarse	fine
Random	0.500	0.500	0.500	0.500	0.500	0.500
Best last	0.694	0.702	0.710	0.702	0.669	0.637
Best ENCOW	**0.782**		**0.831**		**0.669**	
Best diachronic	0.663	0.702	0.651	0.635	0.631	0.633

→ Diachronic setup outperformed by static synchronic baselines
  → Only roughly comparable, very different setup
  → Synchronic compositionality information
  → More data

However: Our focus is on the question of distinct development patterns for the classes.

Results

Features	Accuracy					
	Compound		Modifier		Head	
	coarse	fine	coarse	fine	coarse	fine
Random	0.500	0.500	0.500	0.500	0.500	0.500
F_C	**0.663**	**0.665**	0.595	0.600	**0.631**	**0.633**
F_M	0.585	0.597	**0.649**	**0.629**	0.457	0.455
F_H	0.649	0.647	0.519	0.523	0.627	0.617
F_{MH}	0.637	0.643	0.605	0.624	0.592	0.595
F_{CMH}	0.654	0.644	0.594	0.620	0.570	0.576
P_M	0.629	0.626	0.632	0.606	0.457	0.448
P_H	0.571	0.564	0.502	0.472	0.554	0.550
P_{MH}	0.612	0.597	0.610	0.607	0.538	0.518
$F_{CMH}P_{MH}$	0.619	0.634	0.590	0.608	0.568	0.574

→ Performance significantly above random baseline for most settings

→ Exceptions match intuition
(Head properties not informative for Modifier and vice versa)

Conclusion

- Novel diachronic approach to compositionality
  - Using linguistic information not obtainable through synchronic resources
- Informative, distinct development patterns in properties of high- vs. low-compositionality compounds
- Potential in under-researched area

{maximilian-martin.maurer,
christopher.jenkins,
filip.miletic,
schulte}
@ims.uni-stuttgart.de

Camera-ready paper

References

Roland Schäfer. 2015.

Processing and querying large web corpora with the COW14 architecture. In *Proceedings of Challenges in the Management of Large Corpora 3 (CMLC-3)*, Lancaster.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016.

Diachronic word embeddings reveal statistical laws of semantic change. In *roceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 1489–1501, Berlin, Germany.

Daniel Jurafsky, Alan Bell, Michelle Gregory, and William D. Raymond. 2001.

Michelle Gregory, and William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Joan Bybee and Paul Hopper, editors, *Frequency and the Emergence of Linguistic Structure, Typological Studies in Language,*, pages 229-–254, John Benjamins, Amsterdam/Philadelphia.

Christopher J. Lee. 1990.

Some hypotheses concerning the evolution of polysemous words. In *Journal of Psycholinguistic Research*, 19(4):211-219.

References

Martin Hilpert. 2015.

From *hand-carved* to *computer- based*: Noun-participle compounding and the up- ward strengthening hypothesis. In *Cognitive Linguistics*, 26(1):1–36.

Silvio Cordeiro, Aline Villavicencio, Marco Idiart, and Carlos Ramisch. 2019.

Unsupervised compositionality prediction of nominal compounds. In *Computational Linguistics*, 45(1):1–57.

Reem Alatrash, Dominik Schlechtweg, Jonas Kuhn, and Sabine Schulte im Walde. 2020.

CCOHA: Clean corpus of historical American English. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6958—6966, Marseille, France. European Language Resources Association.

Examples & Compositionality Ranges

Compound	Compositionality rating		
	Modifier	Head	Compound
guinea pig	0.47 ± 0.72	0.47 ± 0.72	0.24 ± 0.56
flea market	0.38 ± 0.81	4.71 ± 0.84	1.52 ± 1.13
pain killer	4.71 ± 0.64	1.33 ± 1.11	2.05 ± 1.36
health insurance	4.53 ± 0.88	4.83 ± 0.58	4.40 ± 1.17

Table: Sample gold standard compounds with compositionality ratings (mean and standard deviation).

	Compositionality					
	Compound		Modifier		Head	
Class	low	high	low	high	low	high
Range	[0.18, 1.61]	[4.20, 5.00]	[0.14, 1.76]	[4.56, 5.00]	[0.00, 2.79]	[4.50, 5.00]

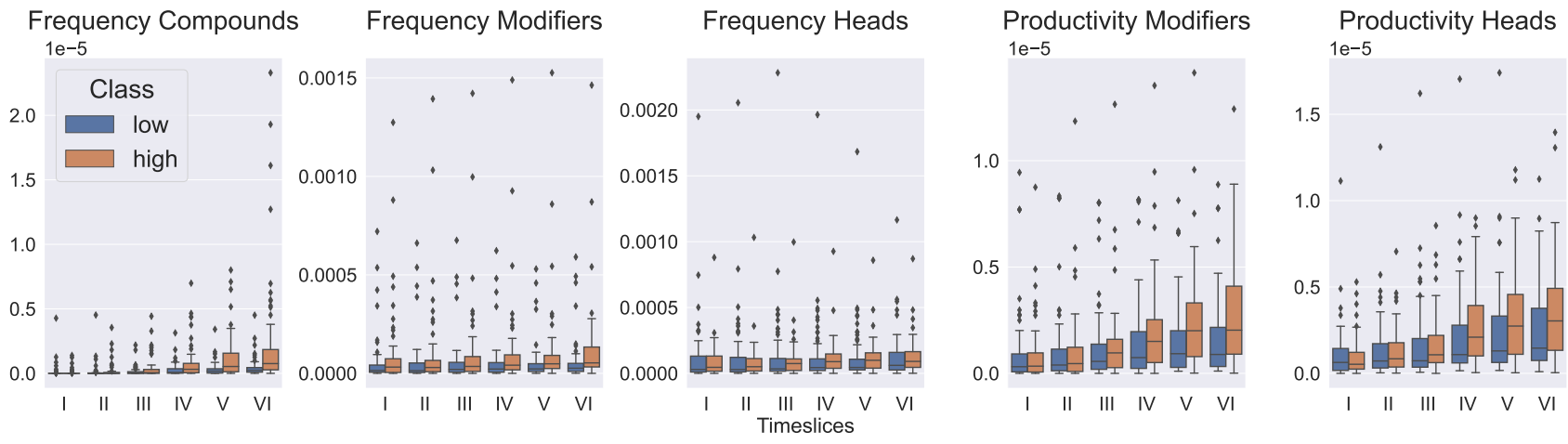
Table: Compositionality ranges for the targets per class and experiment.

Timeslice Sizes

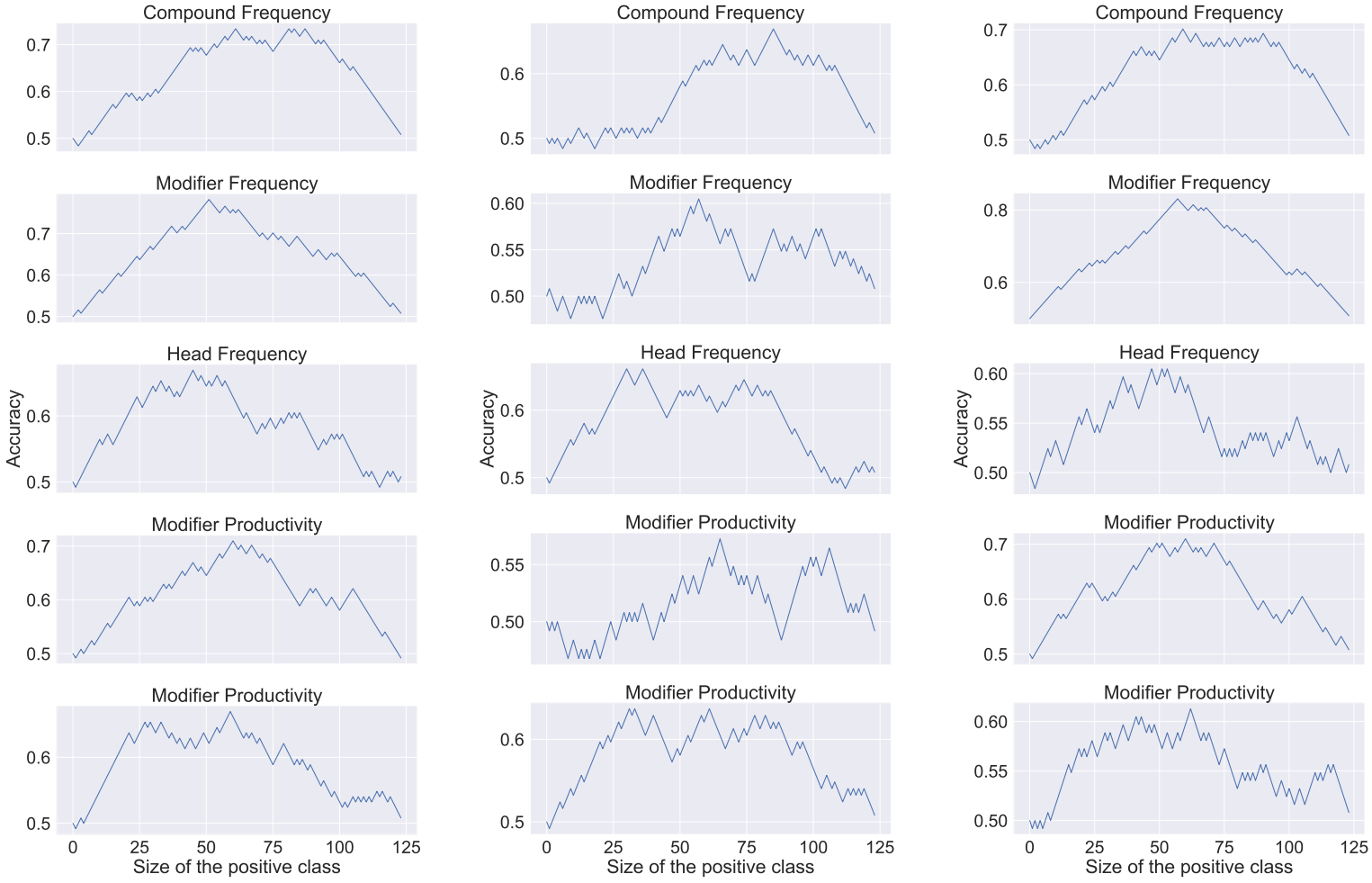
Timeslice	1830s	1840s	1850s	1860s	1870s	1880s
Total_{fine}	16.7	19.4	20.0	20.6	22.6	24.4
Total_{coarse}	56.1			67.6		
Timeslice	1890s	1900s	1910s	1920s	1930s	1940s
Total_{fine}	24.6	26.7	27.7	31.2	30.1	29.9
Total_{coarse}	79.0			91.2		
Timeslice	1950s	1960s	1970s	1980s	1990s	2000s
Total_{fine}	30.3	29.6	29.4	31.3	34.6	36.5
Total_{coarse}	89.3			100.4		

Table: Timeslice sizes for the fine- and coarse-grained timeslices in million tokens.

Developent of Properties over Time per Class



Best N ENCOW


Compound Frequency
Modifier Frequency
Head Frequency
Modifier Productivity
Modifier Productivity
Accuracy
Size of the positive class
0
25
50
75
100
125

Examples Developments Frequency and Productivity