Probing BERT for German Compound Semantics

Filip Miletić • Aaron Schmid • Sabine Schulte im Walde
{filip.miletic, aaron.schmid, schulte}@ims.uni-stuttgart.de

University of Stuttgart
Institute for Natural Language Processing

Motivation

Noun compounds are ubiquitous yet challenging to model
→ variable degrees of compositionality
→ cross-lingual differences in productivity & structures

We replicate a prior probing study of English BERT
(Miletić & Schulte im Walde, 2023)

More challenging scenario of German
→ closed spelling
→ higher compounding productivity
→ higher constituent-level ambiguity

Can BERT capture degrees of noun compound compositionality in German?

How does German BERT compare to prior work?

Approach	Modif.	Head
Our best result	0.332	0.433
Schulte im Walde et al. (2016a) LMI vectors; same data	0.490	0.590
Miletić & Schulte im Walde (2023) same method; English data	0.553	0.645

Best results in our study and related prior work (Spearman's ρ)

Why the performance drop?
→ inherently higher difficulty of the task for German
→ less efficient learning of semantics due to ambiguity
→ dataset differences, e.g. constituent family sets

How do modeled tokens affect predictions?

		mod	head	comp	cont	cls
Modifier	mod		**0.170**	**0.174**	**0.332**	**0.266**
	head	0.170		0.130	0.019	0.024
	comp	0.174	0.130		0.154	0.113
	cont	**0.332**	0.019	0.154		0.123
	cls	0.266	0.024	0.113	0.123	
Head	mod		0.327	0.202	0.178	0.084
	head	**0.327**		0.290	**0.433**	**0.246**
	comp	0.202	0.290		0.318	0.149
	cont	0.178	**0.433**	**0.318**		0.096
	cls	0.084	0.246	0.149	0.096	

Best individual results obtained using direct comparisons of pairs of embeddings for modifier predictions (top) and head predictions (bottom). Bold values are best in a column; shaded values are best overall.

Example: given the compound *Obstsaft*, we obtain predictions comp-mod: cos(*Obstsaft*, *Obst*); comp-head: cos(*Obstsaft*, *Saft*); etc.

Experimental setup

Corpus: DECOW16 → 11.6bn tokens
(Schäfer & Bildhauer, 2012; Schäfer, 2015)

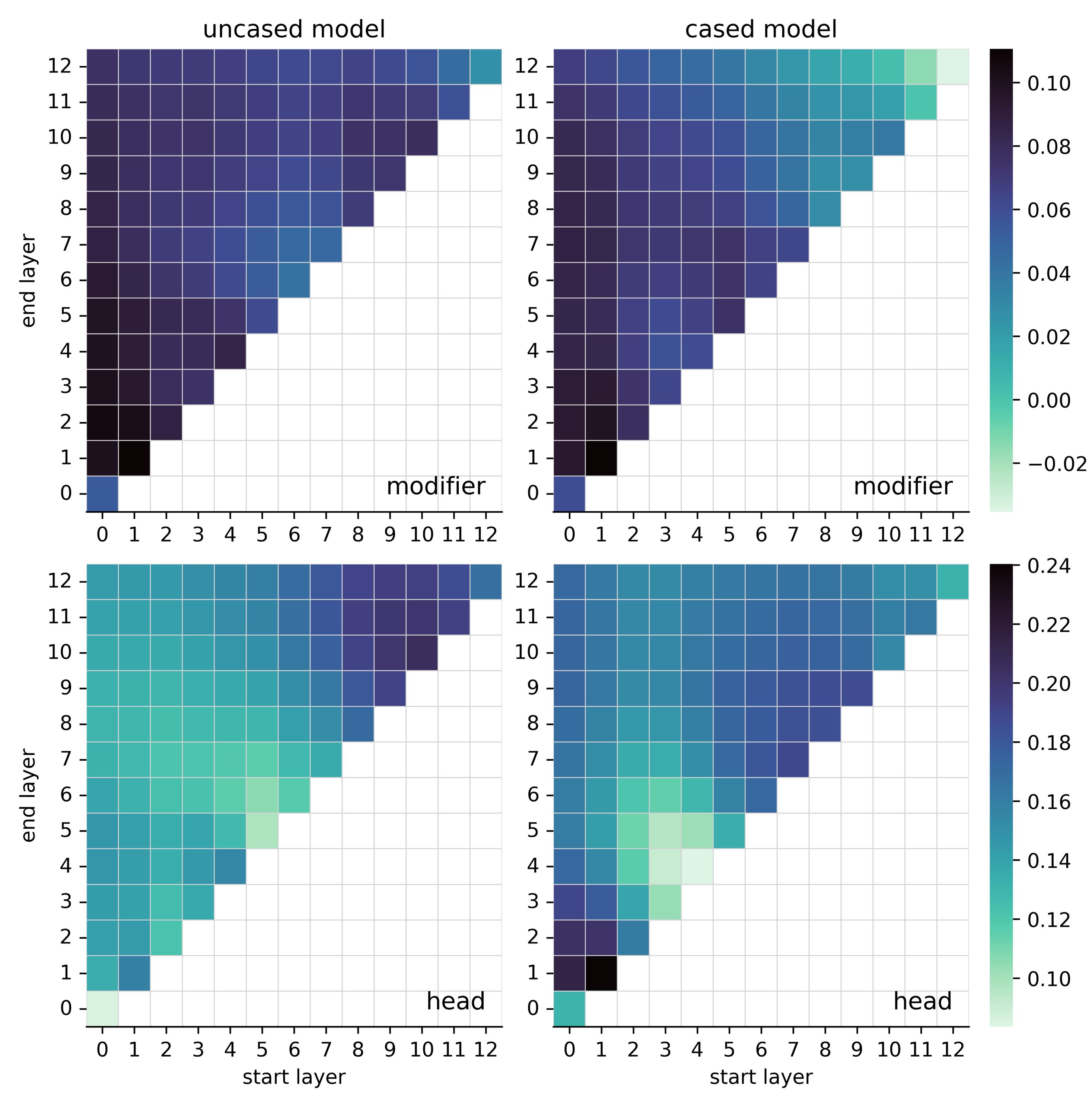
Gold standard: GhoSt-NN → 868 NN compounds
(Schulte im Walde et al., 2016b)

Compound	Modifier		Head	
Erbsensuppe *pea soup*	*Erbse* *pea*	5.3	*Suppe* *soup*	5.3
Eifersucht *jealousy*	*Eifer* *zeal*	2.0	*Sucht* *addiction*	2.1

Sample compounds and ratings (scale 1–6) for modifier and head compositionality

Models: German BERT {cased, uncased}
Tokens: modif, head, comp, cont, cls
Layers: all contiguous spans
Compositionality estimation:
→ direct — pairwise cosine
→ composite — ADD, MULT, COMB

Which layers & models best encode compositionality information?



Mean performance across contiguous spans of layers, defined by the start layer (x-axis) and end layer (y-axis). Left: uncased model; right: cased model. Top: modifier predictions; bottom: head predictions.

References
Miletić, F., & Schulte im Walde, S. (2023). A systematic search for compound semantics in pretrained BERT architectures. In *Proc. EACL*. • **Schäfer**, R. (2015). Processing and querying large web corpora with the COW14 architecture. In *Proc. CMLC-3*. • **Schäfer**, R., & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. In *Proc. LREC*. • **Schulte im Walde**, S., Hätty, A., & Bott, S. (2016a). The role of modifier and head properties in predicting the compositionality of English and German noun-noun compounds: A vector-space perspective. In *Proc. *SEM*. • **Schulte im Walde**, S., Hätty, A., Bott, S, & Khvtisavrishvili, N. (2016b). GhoSt-NN: A representative gold standard of German noun-noun compounds. In *Proc. LREC.*

SwissText 2025
Winterthur