

The German FrameNet-Constructicon

Construction Grammar is a linguistic theory which assumes that a language can be fully described by so called constructions, i.e. form-meaning pairs (Goldberg 1995, Croft 2001). Hence in construction grammar there is no distinction between a lexicon and a grammar, there are only constructions that can be combined to create phrases, sentences and whole texts. The project “German FrameNet-Constructicon” (GFNC) is located in the linguistics department of the institute for German studies. It “researches and documents coherent form-meaning structures of the contemporary German language in the continuum between lexicon and grammar” (<https://framenet-constructicon.hhu.de/>).

The Constructicon

A constructicon contains form-meaning pairs, the building blocks of a language. Such building blocks can be single words like

[tree / ʦ]

or more abstract constructions like a sentence template

[X gives Y sth. / *Transfer of possession*].

Constructions like the above have slots that can be filled with instances of other constructions, e.g. words. The aim of the GFNC is to create a structured repository of constructions for the language German.

The FrameNet

The German FrameNet is based on the theory of frame semantics (Fillmore 1976), which describes how words and their meanings are organized in conceptual structures (frames). Each frame represents a semantic framework motivated by semantically similar words. A lexical unit (LU) and its reading can evoke a specific frame (e.g. “to run” can evoke different frames depending on the context: “run for presidency”, “my nose is running”, “run to the goal”). The German FrameNet serves as a repository of concepts (meanings evoked by a form) that are interconnected in a network.

Construction Mining

There exist different strategies for the identification of meaning bearing forms for the inclusion in a constructicon. One is to look for grammatical patterns in the linguistic literature that can be or have been described as constructions. In the GFNC project, we are also experimenting with a more data-driven approach: construction mining (Barteld and Ziem 2020).

In a first step, form-patterns are extracted from a corpus. We use a corpus annotated with syntactical information to extract patterns that consist of **words** and slots with categorial and functional restrictions in the form of POS tags and **grammatical functions**:

NOUN [die, nmod]

In a second step, these patterns are ranked according to their probability of being meaning-bearing constructions. In first experiments, we have used UIF-PMI (Forsberg et al. 2014) as a measure.

	Frequency	UIF	UIF-PMI	(data from the HDT)
NOUN [die, nmod]	8347	7829	35052.27	
NOUN [ADP, der, nmod]	3447	3243	22260.00	
NOUN [ADP, die, nmod]	3115	3243	20163.54	
NOUN [DET, NOUN der]	3551	2965	19805.49	

These lists are then annotated regarding the construction status of the patterns. In an experiment using the Hamburg Dependency Treebank (HDT) (Foth et al. 2014) patterns including parentheses emerged and led to the inclusion of parenthesis-constructions with different functions into the constructicon.

- ▶ PROPN [(,)] - Abbreviation: *Hamburg Dependency Treebank (HDT)*
<https://framenet-constructicon.hhu.de/constructicon/construction?id=917>
- ▶ PROPN [(,)] - Affiliation: *Frank-Walter Steinmeier (SPD)*
<https://framenet-constructicon.hhu.de/constructicon/construction?id=900>
- ▶ NOUN [(, nummod,)] - Attribution: *1.6 millionen dollar (1.2 millionen euro)*
<https://framenet-constructicon.hhu.de/constructicon/construction?id=590>

Collexeme Prediction

joint work with Tim Feldmüller

Collo-profiles are lists of the most relevant fillers, so called collo-items, for a given slot in a construction (Herbst 2018). These relevant fillers give clues about the meaning of a construction.

Example - The ditransitive construction

Form: [$S_{agent} VO_{recipient} O_{theme}$]

Meaning: ‘X causes Y to have/receive Z’

Collexemes: give, tell, send, offer, ...
(cf. Stefanowitsch and Gries 2003)

The established way of determining collo-items, collexeme analysis (Stefanowitsch and Gries 2003), needs counts of words that appear in a given slot and therefore annotated corpus data.

	Ditransitive	Other constructions	Sum	Table 2 from Gries (2023)
Give	461	699	1160	
Other verbs	574	136,930	137,504	
Sum	1,035	137,629	138,664	

Using a pre-trained language model like BERT (Devlin et al. 2019), we predict probable fillers for a given slot in example sentences following the method for lexical simplification proposed by Wada et al. (2022). With this method it is possible to create collo-profiles for constructions given only example sentences and include also collo-items that do not appear in these sentences (Feldmüller, Barteld, and Ziem in preparation).

Collo-profile for an exemplary construction giving the frequency of an event in the form ‘about once a year’: (only collo-items in bold face appear in the given example sentences)

mindestens, spätestens, wenigstens, höchstens, zumindest, **etwa**

<https://framenet-constructicon.hhu.de/constructicon/constructiondata?id=61>

Bibliography

- Barteld, Fabian and Alexander Ziem (2020). “Construction mining: Identifying construction candidates for the German constructicon”. In: *Belgian Journal of Linguistics* 34, pp. 5–16.
- Croft, William (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press.
- Devlin, Jacob et al. (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Ed. by Jill Burstein, Christy Doran, and Thamar Solorio. Minneapolis, Minnesota: Association for Computational Linguistics, pp. 4171–4186.
- Feldmüller, Tim, Fabian Barteld, and Alexander Ziem (in preparation). “Can BERT predict Fillers For Construction Elements? Adding Collo-Profiles to the German Constructicon”.
- Fillmore, Charles J. (1976). “Frame Semantics and the Nature of Language”. In: *Annals of the New York Academy of Sciences* 280.1, pp. 20–32.
- Forsberg, Markus et al. (2014). “From construction candidates to constructicon entries: An experiment using semi-automatic methods for identifying constructions in corpora”. In: *Constructions and Frames* 6.1, pp. 114–135.
- Foth, Kilian A. et al. (2014). “Because Size Does Matter: The Hamburg Dependency Treebank”. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*. Ed. by Nicoletta Calzolari et al. Reykjavik, Iceland: European Language Resources Association (ELRA), pp. 2326–2333.
- Goldberg, Adele E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
- Gries, Stefan T. (2023). “Collostructional Methods”. In: *The Encyclopedia of Applied Linguistics*. John Wiley & Sons, Ltd, pp. 1–6.
- Herbst, Thomas (2018). “Is language a Collostruction? – A Proposal for Looking at Collocations, Valency, Argument Structure and Other Constructions”. In: *Lexical Collocation Analysis: Advances and Applications*. Ed. by Pascual Cantos-Gómez and Moisés Almela-Sánchez. Springer, pp. 1–22.
- Stefanowitsch, Anatol and Stefan T. Gries (2003). “Collostructions: Investigating the interaction of words and constructions”. In: *International Journal of Corpus Linguistics* 8.2, pp. 209–243.
- Wada, Takashi et al. (2022). “Unsupervised Lexical Substitution with Decontextualised Embeddings”. In: *Proceedings of the 29th International Conference on Computational Linguistics*. Ed. by Nicoletta Calzolari et al. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, pp. 4172–4185.