

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 / 🌳]

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.

Enriching Frames Using Lexical Substitution

To expand frames with new LUs, a substitution method utilizing the BERT language model is applied. Existing LUs of a frame are extracted along with their example sentences from the German FrameNet database. BERT then generates contextually appropriate substitutes based on the LUs within these example sentences, ensuring they preserve the original meaning. These substitutes are subsequently annotated and evaluated. This method, based on lexical substitution, aims to generate lexical representations of frames as described in Anwar et al. (2020).

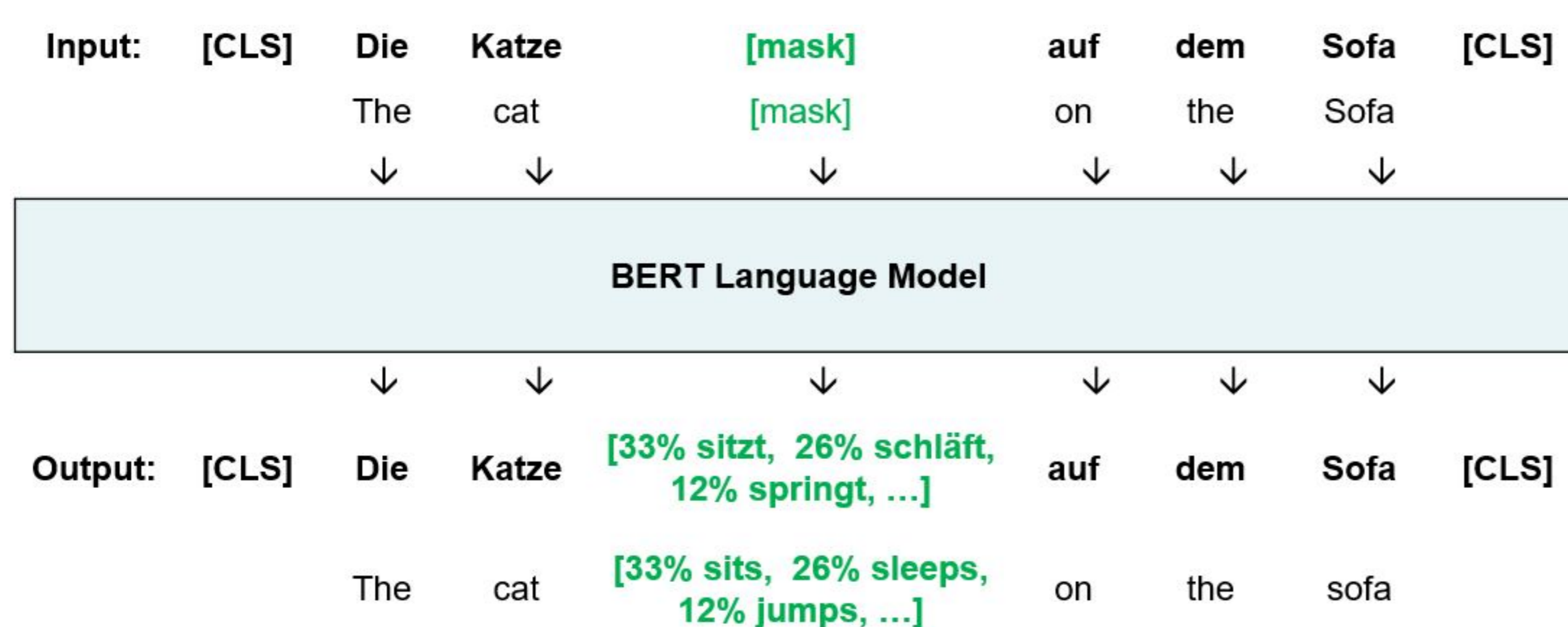


Figure 2: Predicting a masked word in context

Result: For 2104 example sentences, 10 substitutes each were annotated. Initially, there were 4269 LUs. By the end of the process, the German FrameNet comprised 5078 LUs. Thus, 809 new and useful LUs were automatically generated, contributing to the expansion of the German FrameNet.

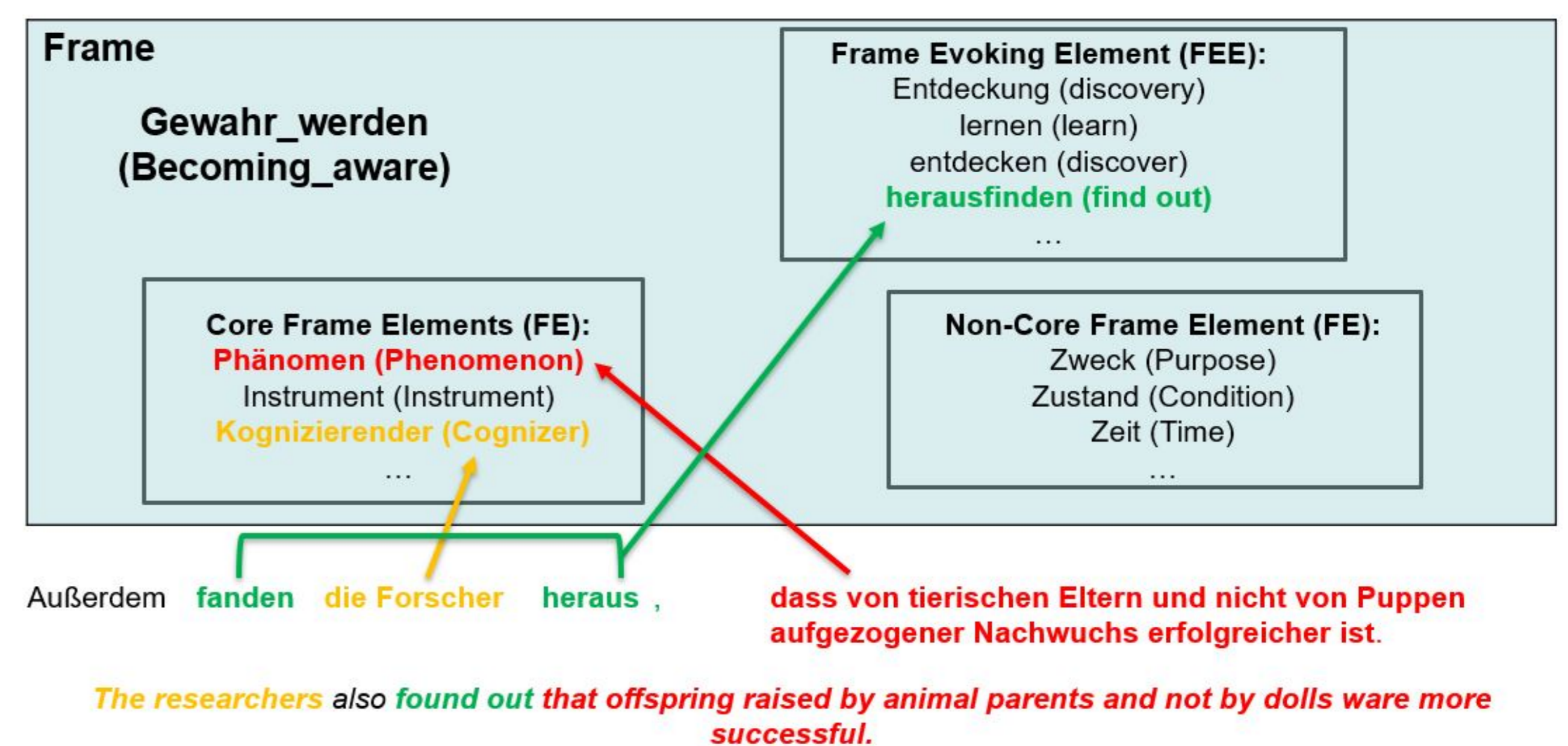


Figure 1: LU evoking a Frame with Frame Elements

Procedure: In the case of the German FrameNet, sentences that contain an instance of the LUs of a frame were taken and the frame evoking element (FEE) - usually verbs - in the sentence was masked. The predicted substitutes from BERT for the gap were then evaluated based on their adequacy for evoking the specific frame.

FRAME_NAME	FRAME_ID	TOKENIZED_SENTENCE	LU	SUB_1	SUB_2	...
Bewegung_in_eine_Richtung	995	Die Frau fiel über die Schuhe ihres Sohnes	fallen	stürzen	springen	steigen, fliegen, sinken, ...
Handel_verkaufen	352	Polen verkauft Gänse nach Deutschland	verkaufen	liefern	produzieren	vermitteln, schicken, ziehen, ...

Figure 3: Generated substitutes for example sentences

Obtaining Natural Example Sentences for New LUs

Based on BERT and the existing German FrameNet data, binary classifiers were trained for each frame to predict how well a sentence fits a given frame. For the newly generated LUs, sentences containing these LUs were searched within the Wikipedia corpus. The corresponding classifier then evaluated these sentences, classifying them according to their suitability for the frame. The top example sentences were selected accordingly.

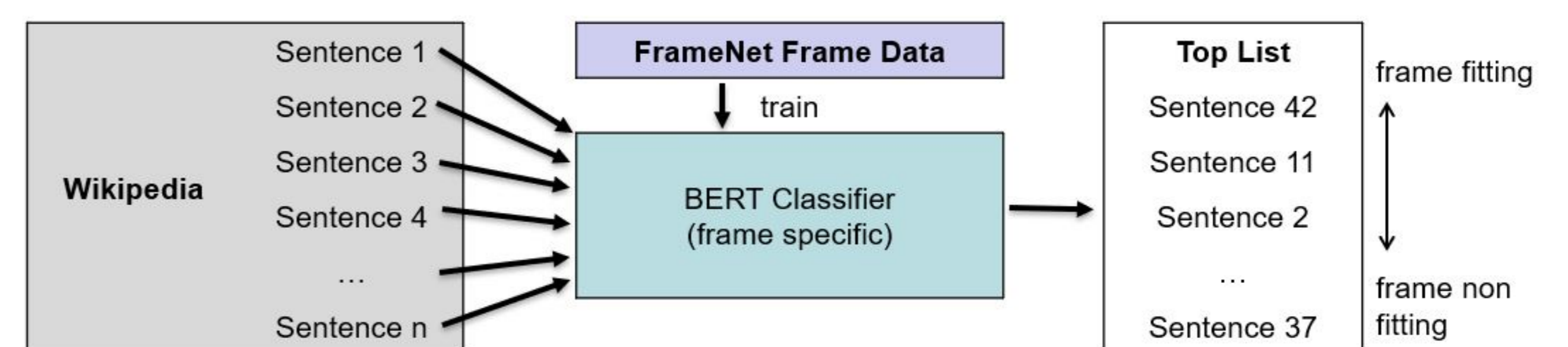


Figure 4: Classifying frame-fitting sentences for a new LUs

Bibliography

- Anwar, Saba et al. (2020). “Generating Lexical Representations of Frames using Lexical Substitution”. In: *Proceedings of the Probability and Meaning Conference (PaM 2020)*. Ed. by Christine Howes et al. Gothenburg: Association for Computational Linguistics, pp. 95–103.
- Croft, William (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press.
- 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.
- Goldberg, Adele E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.