

**Supervisor:** Leonie Weißweiler

**Examiner:** Hinrich Schütze

**BSc, MSc, Open:** Open

**Title:** Unsupervised Induction of Construction Grammars

**Summary:**

Construction Grammar is a family of theories from Cognitive Linguistics in which constructions are the building blocks of language. Constructions are learned combinations of linguistic patterns with meaning. Constructions occur on a spectrum from fixed, such as individual words or morphemes, to very open, such as the passive voice. In the middle of this spectrum, one might find idioms and fixed expressions.

Pioneering work by Jonathan Dunn (Dunn, 2017) <sup>1</sup> has proposed an algorithm to learn these constructions for any given language from a sufficiently large corpus. The algorithm uses complex co-occurrence statistics and association measures to extract the constructions from a corpus completely without labels, and the paper measures its success by analysing the coverage of the resulting grammar for the corpus.

There are several avenues to explore for a thesis. As the code for the paper is available<sup>2</sup>, both to run the pretrained grammars and to train new ones, one possibility would be to compare the results for grammars trained on corpora from different domains and analyse reasons for the difficulties. Another valuable contribution could be training grammars on new languages and making the result available as a resource for cross-lingual research on Construction Grammars. On the more challenging end, there are many possibilities to explore expanding and improving the algorithm, for example with different association measures, or by incorporating other resources or taggers.

**Prerequisites (if any):** Python

---

<sup>1</sup> Dunn, J. (2017). Computational learning of construction grammars. *Language and cognition*, 9(2), 254-292.

<sup>2</sup> [c2xg/Parser.py at master · jonathandunn/c2xg \(github.com\)](https://github.com/jonathandunn/c2xg/blob/master/c2xg/Parser.py)