Streamlining Formal Construction Grammar[1]

Vanja Sophie Cangalovic, Robert Porzel and John A. Bateman
Departments of Computer Science & Linguistics, University of Bremen

Streaming Construction Grammar (SCG) is both a Construction Grammar formalism (Fillmore, 1988; Goldberg, 1995) and an implemented parser used in simulation-based natural language understanding in the context of [removed for anonymity]. It combines the focus on constructions as mental simulation-parameterising operations and the tight integration with external knowledge bases from Embodied Construction Grammar (ECG; Feldman et al., 2009) with the computational model of unification and merging from Fluid Construction Grammar (FCG; Steels et al., 2006). SCG uniquely simplifies data structures, mechanisms, and semantic output through the pervasive use of triples, a.k.a. [unit feature value] 3-tuples, easing formalisation, implementation, and integration with robotic and knowledge representation systems.

An SCG grammar is responsible for transforming underspecified natural language instructions from the kitchen domain into interpretations of an ontological model (Beßler et al., 2020). The schemas evoked by the input sentence parameterise a subsequent simulation which is able to determine unknown parameter values and the overall plausibility of the instruction interpretation.

Conversely, knowledge from an ontology, a simulation, or robotic sensors can be used to guide the search process, enable context-aware constructions, or to resolve coreferences via both syntactic heuristics as well as semantic constraints (Raghuram et al., 2017). In constructions it enables evoking unspecified parameters (to be inferred by later processing steps) and facilitates the processing of syntactically ambiguous instructions (e.g. by distinguishing CAUSEDMOTION from PROPERTYRESULTATIVE). Continuously streaming this information into the grammar is what gives SCG its name.

This connection is realised through a new query operator executed between the match and merge operations of FCG, which results in new bindings between the query variables and values found in the database, e.g. as a success criterion they express role filler type constraints like those in ECG.

Although the loose structure of triples naively increases computational complexity compared to classical attribute-value-matrixes, this is mediated by optimisation techniques adapted from database theory. For instance, a Datalog-inspired semi-naive evaluation strategy was implemented, and SAT-based constraint solving and chart-like parsing similar to the RETE-algorithm are currently being worked on, with the simpler model allowing for accelerated experimentation and exploration.

The SCG system also provides a notebook-based playground for exploring constructions and parsing strategies, external knowledge bases, and information-providing systems.

References

Beßler, D., Porzel, R., Pomarlan, M., Vyas, A., Höffner, S., Beetz, M., Malaka, R. & Bateman, J. (2020) Foundations of the Socio-physical Model of Activities (SOMA) for Autonomous Robotic Agents. arXiv 2011.11972.

Feldman, J., Dodge E., & Bryant J. (2009). A Neural Theory of Language and Embodied Construction Grammar. The Oxford Handbook of Linguistic Analysis, pp. 111-138.

Fillmore, C. (1988). The Mechanisms of "Construction Grammar". Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society, pp. 35-55.

Goldberg, A. E. (1995). Constructions: A Construction Grammar Approach to Argument Structure. Cognitive Theory of Language and Culture Series.

Raghuram, V., Trott, S., Shen, K., Goldberg, E.M., & Oderberg, S. (2017). Semantically-Driven Coreference Resolution with Embodied Construction Grammar. *AAAI Spring Symposia*.

Steels, L. & De Beule, J. (2006). Unify and Merge in Fluid Construction Grammar. Symbol Grounding and Beyond, pp. 197-223.

[1] The project "MUHAI" leading to this application has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 951846.