A bidirectional computational construction grammar for German exploiting syntactico-semantic pattern finding on a learners' corpus

Veronica Juliana Schmalz, FWO fellow at KU Leuven, veronicajuliana.schmalz@kuleuven.be

Keywords: Fluid Construction Grammar, usage-based language acquisition, German grammar

For over thirty years, early language acquisition has been studied through usage-based approaches. Empirical studies have revealed that children can incrementally learn a grammar without any previous knowledge exploiting only communicative interactions with language users through intention reading and pattern finding cognitive capacities (cf. Tomasello 1992, 2003, Goldberg 1995). More recently, usage-based constructionist approaches have focused on L2 acquisition (Ellis 2013, De Knop 2016), highlighting the similarities with L1 acquisition. According to them, L2 acquisition would be a process of "construction and reconstruction" (Ellis 2013, p.2) of linguistic knowledge: learners need to continuously monitor the different levels of abstractions and networks that their L1 and L2 constructions form with each other. However, closely analyzing these aspects with construction grammar (CxG) paradigms resulted rather intractable. Thanks to novel technologies and computational methodologies (Van Eecke 2018, Van Eecke and Beuls 2018) we can now investigate in detail the processes of CxG acquisition.

Despite the rise in interest and research on CxG acquisition, current knowledge appears to be rather limited and fragmented. First, existing models require either previous linguistic knowledge or extensively annotated data (Dominey and Boucher 2005, Chang 2008). Second, the different adopted formalisms and the restricted complexity of the methods -or data types- render the models detached from human-like language capabilities (Spranger and Steels 2015). In this regard, our aim is to **computationally model the acquisition of usage-based CxGs** in an efficient and traceable manner.

We present our model to monitor the co-acquisition of modular constructions and the emergence of grammatical categories in a usage-based manner, exploiting the formalism of FCG or Fluid Construction Grammar (Steels 2011) on an augmented corpus of German learners' utterances (Baten and Cornillie 2019) annotated with Abstract Meaning Representations (Banarescu et al. 2013). We focus on L2 data since CxGs have, indeed, been regarded as a suitable ICALL framework by Schulze and Penner (2008). However, to date, no computational model can represent in detail the evolution of learners' CxG repertoire in comprehension and production.

Our methodology stems from Nevens et al. 2022 and Doumen et al. 2023 and employs generalization and specialization mechanisms to detect patterns in utterances. In this way, we can automatically enable the acquisition of constructions through syntactico-semantic pattern finding operators and monitor their interaction through the emergence of categorial networks. With our experiment on this corpus we aim to provide a proof of concept showing that with a limited set of examples (100 utterances), different types of progressively more modular constructions can be acquired (ex. ?x-*gibt*-*?y*-*?z*-*cxn*). These constructions encapsulate both general grammatical aspects, such as verbal arguments in transitive, ditransitive and intransitive utterances, and more detailed aspects such as the cases and nominal agreement in German.

In conclusion, our research is intended to serve a twofold purpose: first, it provides a more in-depth understanding about the automatic acquisition of modular constructions and the emergence of grammatical categories from simple utterances. Second, it offers the possibility of bootstrapping a grammar from any semantically annotated corpus. On top, this grammar is bidirectional, applicable for both comprehension and production, as well as traceable in detail. Through this type of research, we aim to pave the way for truly intelligent systems with human-like language capabilities, applicable in different domains, including smart tutoring, conversational agents, and visual question answering systems.

References

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., & Schneider, N. (2013). Abstract meaning representation for SemBanking. In A. Pareja-Lora, M. Liakata & S. Dipper (Eds.), Proceedings of the 7th linguistic annotation workshop and interoperability with discourse (pp. 178-186). Association for Computational Linguistics.

Baten, K., & Cornillie, F. (2019). Elicited imitation as a window into developmental stages. Journal of the European Second Language Association, 3(1), 23–34.

Chang, N. (2008). Constructing grammar: A computational model of the emergence of early constructions. PhD thesis, University of California, Berkeley, Berkeley, CA.

De Knop, S., & Gilquin, G. (Eds.). (2016). *Applied construction grammar* (Vol. 10, No. 9783110458268). Boston/Berlin: De Gruyter Mouton.

Dominey, P. F. & Boucher, J.-D. (2005). Learning to talk about events from narrated video in a construction grammar framework. Artificial Intelligence, 167(1):31–61.

Doumen, J., Beuls, K., & Van Eecke, P. (2023). Modelling language acquisition through syntactico-semantic pattern finding. In *The 17th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics.

Ellis, N. C. (2013). Second language acquisition. In Oxford Handbook of Construction Grammar (pp. 365-378), G. Trousdale & T. Hoffmann (Eds.). Oxford: Oxford University Press.

Goldberg, A. (1995). Constructions: A Construction Grammar Approach to Argument Structure. University of Chicago Press, Chicago.

Nevens, J., Doumen, J., Van Eecke, P., & Beuls, K. (2022, October). Language Acquisition through Intention Reading and Pattern Finding. In *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 15-25).

Spranger, M. & Steels, L. (2015). Co-acquisition of syntax and semantics: an investigation in spatial language. In Yang, Q. & Wooldridge, M., editors, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, pages 1909–1915, Palo Alto, Ca. AAAI Press.

Steels, L. (2011). Design patterns in fluid construction grammar. John Benjamins.

Tomasello, M. (1992). The social bases of language acquisition. Social development, 1(1), 67-87.

Tomasello, M. (2003). Constructing a language: A usage-based theory of language acquisition. Harvard University Press.

Van Eecke, P. (2018). Generalisation and Specialisation Operators for Computational Construction Grammar and their Application in Evolutionary Linguistics Research. PhD thesis, Vrije Universiteit Brussel, VUB Press, Brussels.

Van Eecke, P. & Beuls, K. (2018). Exploring the creative potential of computational construction grammar. Zeitschrift ür Anglistik und Amerikanistik, 66(3), 341-355.