

A linear, incremental take on Construction Grammar: introducing the FACETS annotation framework for keylogging data

Annotating sentences with Construction Grammar has proven a challenging theoretical issue for this framework (Torrent et al. 2014, Muralidaran & Misra Sharma 2016). Any typical sentence indeed involves constructions of various levels of schematicity, corresponding to the multiplicity of levels of a vertically organized hierarchic network that relates the most abstract constructions to the most specified constructions (Langacker 2001, Boas & Ziem 2018, Herbst & Hoffmann 2018). These annotations issues become all the more stringent when it comes to unconventional data, such as oral data (Fried & Östman 2005, Gilquin 2022) or real-time writing data as offered by keylogging software (Gilquin 2020). In the latter, the whole of the writing process is recorded as writers type their texts, including pauses, revisions, and non-linearities (Conijn et al. 2019, Conijn et al. 2024). To annotate these data meaningfully, one therefore needs to rely on an incremental, linear approach (Frank et al. 2012), in line with other usage-based perspectives on language such as the Linear Unit Grammar (Sinclair & Mauranen 2006). The goal shifts from inventorying all possible structures writers draw from to produce specific utterances, to pinpointing the iterative operations the writer is performing to transition from the production of a token to another, highlighting the choices made on argument structure and on the lexical units that fill them (Christiansen & Chater 2016).

Here we exploit keylogging data from L2 learners of English, writing argumentative essays (Anonym). In an L2 context, writers tend to revise more, making explicit a wider array of cognitive operations insofar as revisions highlight a decision point in which communicative choices need to be made. The keylogging data records all time stamps of all keystrokes underlying the textual production of the essays, including deletion (pressing of the “BACK” key). Our annotation framework, called FACETS, defines the following six operations (the first letters of each forming the acronym). In the examples below, the deleted part is displayed in parenthesis.

- Treading

The writer starts a new utterance, possibly triggered by a discourse structuring marker or a subordinating conjunction, therefore answering a specific discourse need that has been set prior.

(We all remember all the) > We learn from experience

- Structuring

The writer fleshes out a schema of the utterance (i.e. it moves down the vertical hierarchy of constructions), either at the argument structure level or at the phrasal level.

child's parent would obviously want (to) > the criminal to go in prison

- Filling

The writer performs a lexical choice over a fully specified syntactic schema.

raising a crucial (distinction) > element

- Commenting

The writer adds optional elements to a schema.

this means that (punis) > hard punishment

- Appending

The writer uses a coordinating conjunction to duplicate a schema's structure while staying in the same utterance.

chances at finding a job and getting some stability(.) > and balance in life.

- Ending

The writer decides that the utterance is now complete.

Money was almost designed to be stolen by those who cannot control their greed(.) > .

Our contribution will flesh out these operations by extracting a large number of revision events from the corpus, evidencing both the structures that are holistically handled to produce and revise utterances, and the key decisions that writers make in their use of the language.

References

- Boas, H. C., & Ziem, A. (2018). Constructing a constructicon for German: Empirical, theoretical, and methodological issues. In Lyngfelt, B., Boren, L., Ohara, K., Torrent, T. T. (eds.), *Constructicography* (pp. 183-228). John Benjamins.
- Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and brain sciences*, 39, e62.
- Conijn, R., Dux Speltz, E., & Chukharev-Hudilainen, E. (2024). Automated extraction of revision events from keystroke data. *Reading and Writing*, 37(2), 483-508.
- Conijn, R., Roeser, J., & Van Zaanen, M. (2019). Understanding the keystroke log: The effect of writing task on keystroke features. *Reading and Writing*, 32(9), 2353-2374.
- Frank, S. L., Bod, R., & Christiansen, M. H. (2012). How hierarchical is language use?. *Proceedings of the Royal Society B: Biological Sciences*, 279(1747), 4522-4531.
- Fried, M., & Östman, J. O. (2005). Construction Grammar and spoken language: The case of pragmatic particles. *Journal of pragmatics*, 37(11), 1752-1778.
- Gilquin, G. (2020). In search of constructions in writing process data. *Belgian Journal of Linguistics*, 34(1), 99-109.
- Gilquin, G. (2022). Constructing learner speech: On the use of spoken data in Applied Construction Grammar. In H. Boas (ed.) *Directions for Pedagogical Construction Grammar: Learning and Teaching (with) Constructions* (pp. 73-96). Berlin: de Gruyter
- Herbst, T., & Hoffmann, T. (2018). Construction grammar for students: A constructionist approach to syntactic analysis (CASA). *Yearbook of the German Cognitive Linguistics Association*, 6(1), 197-218.
- Langacker, R. (2001). Discourse in Cognitive Grammar. *Cognitive Linguistics*, 12(2), 143-188. <https://doi.org/10.1515/cogl.12.2.143>

Muralidaran, V., & Misra Sharma, D. (2016, April). Construction grammar based annotation framework for parsing Tamil. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 378-396). Cham: Springer International Publishing.

Sinclair, J. M., & Mauranen, A. (2006). Linear unit grammar.

Torrent, T. T., Lage, L. M., Sampaio, T. F., da Silva Tavares, T., & da Silva Matos, E. E. (2014). Revisiting border conflicts between framenet and construction grammar: Annotation policies for the brazilian portuguese constructicon. *Constructions and Frames*, 6(1), 34-51.