

Schemas, chunks and everything between: Evidence from Germanic and Romance.

Olaf Mikkelsen^{1,2} & Dylan Glynn^{1,2}

¹University of Paris 8, ²Adam Mickiewicz University of Poznan

Keywords: future tense, construction grammar, schematicity, corpus linguistics, statistical modelling

The typological literature on European languages (Fleischman 1982, Bybee, Perkins & Pagliuca 1994, Dahl 2000) points to a fundamental divide between intention-based futures on the one hand, and prediction-based futures on the other. The former would typically be used to describe future situations that are intentional, temporally close, psychologically certain, and unconditional; the latter situations that are non-intentional, temporally remote, psychologically uncertain, and conditional. Examples of the first would be futures derived from movement verbs (*going*), examples of the latter would be Romance morphological futures or Germanic *will*. In this talk, we look at the future alternations in English, Norwegian, French and Spanish from a usage-based perspective and ask whether the encoding is best explained in terms of schematic generalisations, (partially or fully) lexically specific chunks, or a combination of the two strategies (Dabrowska 2020). To answer this question, a three-step analysis will be carried out. First, a distinctive collexeme analysis (Gries & Stefanowitsch 2004) is used to identify significant chunks; second, behavioural profile analysis (Divjak & Gries 2006) is used as a proxy for schematic meaning differences; third, a quantitative and qualitative assessment is used to evaluate the interactions between the two first.

To assure the comparability of the data, it is extracted from corpora consisting of online personal blogs for each language. A total of 400 000 future constructions are annotated for subject/grammatical person and lexical verb to produce the input for the collexeme analysis. This makes it possible to go beyond a simple list of verbs and produces a more fine-grained (and cognitively plausible) output of the type *NP (not) going to happen*, *(what) are you going to do*, or *we'll see*. From the complete dataset, a subsample of 2200 items is then manually annotated for a series of usage features based on the abovementioned semantic dimensions (intentionality, temporal proximity, speaker certainty and conditionality). To assure replicability of the results, every example is annotated blindly by two annotators for every language. Multifactorial statistics is then used to model the correlation between each future construction and the semantic variables. Separate analyses are conducted for each language. Finally, the association score for each pattern (subject/grammatical person + cx + verb) is added as a predictor in the statistical model as a way of measuring the relative contribution of the subject/grammatical person and verb to the schemas.

In conclusion, this talk looks at differences and similarities between etymologically similar constructions across related languages at two different levels of abstractions (chunks and schemas). It asks (i) whether we will find significant differences between similar constructions, such as English *will* + inf and Norwegian *vil* + inf, or French *aller* + inf and Spanish *ir a* + inf; (ii) whether the results will be coherent between the two levels of abstractions; and finally (iii) how much the chunks influence the schemas.

References

- Bybee, Joan, Revere Dale Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Dabrowska, Ewa. 2020. Language as a phenomenon of the third kind. *Cognitive Linguistics* 31(2): 213–229.
- Dahl, Östen. 2000. *Tense and Aspect in the Languages of Europe*. Berlin: Mouton de Gruyter.
- Divjak, Dagmar & Stefan Th. Gries. 2006. Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1): 23–60.
- Fleischman, Suzanne. 1982. *The future in thought and language. Diachronic evidence from Romance*. Cambridge: Cambridge University Press.
- Goldberg, Adele. 2019. *Explain me this. Creativity, Competition and the Partial Productivity of Constructions*. Princeton: Princeton University Press.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1): 97–129.
- Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization and entrenchment*. Oxford: Oxford University Press.