

Modeling the variety of constructions with morphological causatives metagrammatically: a first sketch

Valeria Generalova
Heinrich Heine University of Dusseldorf*

A metagrammar, in a broad sense, is a way of describing a grammar of a language. Algorithms for producing formal linguistic descriptions have been developed within individual theories, but recently Crabbé et al. 2013 suggested a metagrammatical formalism for encoding syntactic representations of any tree-based grammar. Later, Lichte and Petitjean 2015 integrated semantic representations in the form of frames into it. XMG is a fully declarative language with a rich and well-developed system of typed features. Each description is realized as a class with several dimensions: usually syntactic and semantic, but also morphological and pragmatic. Classes can inherit information from each other, reflecting a system of generic and specific descriptions.

So far, XMG has been used for describing individual languages (Duchier et al. 2012; Burkhardt, Lichte, and Kallmeyer 2017 *inter alia*). It has also been proven to work well with irregular pieces of grammar, e. g. multi-word expressions (Lichte, Petitjean, et al. 2018) or Russian verbal prefixes (Zinova 2017). However, XMG might be a powerful instrument for cross-linguistic studies. The idea of using the same comparative concepts as features and sharing information between classes that describe individual languages is in line with general trends in linguistic typology.

In our paper, we suggest a first sketch of a cross-linguistic metagrammar. It describes constructions with morphological causatives with specific emphasis on causatives derived from transitive base verbs. Causatives are present in the majority of languages, which gives our metagrammar much potential for development. Constructions with morphological causatives usually demonstrate regular syntactic and semantic patterns that are easy to model using a formal description. The current version of our metagrammar was based on and thus intended to describe five languages: Bashkir, Finnish, Halkomelem, Kalmyk, Trumai. This sample has no intention to be typologically representative but aims to reflect some major differences in language structures. One can note that we chose languages with different alignment types, different word orders, and different marking strategies in causative constructions. We also chose genetically close Bashkir and Kalmyk languages to ensure that our metagrammar describes some precise language differences. In our paper, we will show that new languages are easy to interpolate in the metagrammar designed as described.

We present a novel organization of a metagrammar oriented on cross-linguistic studies. The idea is that it is built of two independent parts, and the description of an individual construction in a given language results from combining a *construction class* and a *language plugin*.

*This work is part of the [TreeGraSP](#) project funded by the ERC Consolidator Grant awarded to Prof. Laura Kallmeyer.

Construction classes describe constructions as a theoretical construct. They account for all necessary features (e.g., word order, case of each NP, transitivity, etc.), but do not specify their values. However, the restrictions on feature values can apply. For example, a causative construction can have three arguments iff the base verb is transitive. Our metagrammar allows 3-argument constructions with transitive bases and forbids them with intransitive bases. Also, construction classes account for all possible word orders without giving preference to any of them.

This idea of accounting for a feature without giving it a value is crucial for our approach. Thus, we declare classes with equal sets of features and list options depending on their values through disjunction. Classes inherit one from another in such a way that only one disjunction is introduced per class. Needless to say that the formulation of this set of relevant features is impossible without a thorough typological background. In our paper, we provide a detailed discussion of the way how these features have been selected.

As a result, a construction class is able to describe any construction in question in any language of the sample. In other words, only one class is needed to parse any of the diverse constructions. When it comes to generating sentences or proving their grammaticality for each particular language, these general classes are not sufficient. In our architecture, it is easy to proceed towards descriptions of individual languages using general construction classes. To achieve that, the second part of the architecture serves, i.e., language plugins. A language plugin is a list of features with specified values. These features describe the system of a language and help to select varieties of general construction classes that are relevant for this language. For example, Bashkir and Kalmyk are verb-final languages, which is specified in their plugins. Subsequently, only verb-final constructions are generated for them or accepted as grammatical while parsing. We give more information about interactions between construction classes and language plugins in our paper. In particular, we show some cases when several causative constructions coexist in one language. We also discuss interactions between language-specific features (e.g., word order) and construction-specific features (e.g., transitivity of a predicate).

We claim that a metagrammar designed in this way could describe constructions in all languages once two principal conditions are met. Firstly, construction classes must account for an exhaustive set of options encountered in world languages (see discussion about syntactic patterns of causative constructions in Dixon 2000, pp. 50–55). Secondly, features have to be formulated as comparative concepts in order to have all language plugins in a unified form (see discussion in Haspelmath 2010). Our paper presents work in progress; therefore, we are not ready to report this done. Nevertheless, we show that our metagrammar can deal with languages that were not used for its design. Namely, we show how it describes causative constructions in Hebrew and Nivkh.

Another important finding of our architecture is an extensive system of class inheritance. It means that in both construction classes and language plugins, information can be received from or transmitted to another class. This system allows us to build some sort of an invariant of a causative construction in general, as well as a causative pattern common to all accusative languages, etc. Thanks to the explicit declaration of inheritance relations, we can visualize the variety of causative constructions as a tree and observe how close or how different the elements are. This structure can bring new insights concerning conceptualizing the typology of constructions, although the material presented in our paper is too small to give more than a promise.

The outcome of our work is twofold. Firstly, we prove that XMG is a useful instrument for cross-linguistic studies. In combination with other methods and tools, it could serve to create various resources for typologists and broader audience. Secondly, we enter the debate about

parameters relevant for studying causative constructions and their applicability to different language systems and present our view of this in a formalized way.

References

- Burkhardt, Benjamin, Timm Lichte, and Laura Kallmeyer (2017). “Depictives in English: An LTAG Approach”. In: *Proceedings of the 13th International Workshop on Tree Adjoining Grammars and Related Formalisms*, pp. 21–30.
- Crabbé, Benoit et al. (2013). “XMG: extensible metagrammar”. In: *Computational Linguistics* 39.3, pp. 591–629.
- Dixon, Robert M W (2000). “A typology of causatives: form, syntax and meaning”. In: *Changing valency: Case studies in transitivity*. Ed. by Robert M W Dixon and Alexandra Y Aikhenvald. Cambridge: Cambridge University Press. Chap. 2, pp. 30–83.
- Duchier, Denys et al. (2012). “Describing morphologically-rich languages using metagrammars: a look at verbs in ikota”. In: *Language Technology for Normalisation of Less-Resourced Languages*, p. 55.
- Haspelmath, Martin (2010). “Comparative concepts and descriptive categories in crosslinguistic studies”. In: *Language* 86.3, pp. 663–687.
- Lichte, Timm and Simon Petitjean (2015). “Implementing semantic frames as typed feature structures with XMG”. In: *Journal of Language Modelling* 3.1, pp. 185–228.
- Lichte, Timm, Simon Petitjean, et al. (2018). “Lexical encoding formats for multi-word expressions. The challenge of “irregular” regularities”. In: *Representation and Parsing of Multiword Expressions*. Language Science Press, pp. 41–72.
- Zinova, Yulia (2017). “Modelling Derivational Morphology: A Case of Prefix Stacking in Russian”. In: *International Conference on Formal Grammar*. Springer, pp. 125–141.