

## Universal Grammar Today

Aarne Ranta

Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg

Universal grammar has a long history ranging from mediaeval ideas about the grammar being "the same" in all languages to Chomsky's theory about innate language faculty in humans. This talk is on the lines of the mediaeval idea: we want to share grammatical structures and rules across languages. The context of this work is computational: we want to test and exploit our theories in actual processing of different languages. We make no *a priori* claims that they will apply to all possible human languages, but rather add one language at a time and see how far we can get.

The task of a formal grammar is, traditionally, to define the set of all and only the grammatical expressions of a language. This ideal has largely been abandoned in computational linguistics, because it is too hard to specify what precisely is grammatical in natural language. Instead, we approximate languages either from above or from below. When analysing a language, we approximate it from above: we want to make sure not to miss anything that belongs to the language, and it does not matter if the grammar used for analysis is "overgenerating". When generating a language, we approximate it from below: we want to make sure that the output is grammatically correct, but it does not matter if the grammar fails to generate some parts of the language.

We will explain two contemporary approaches that use common structures for different languages. In the analysis direction, UD (Universal Dependencies) has been used for over 150 languages. In the direction of generation, GF (Grammatical Framework) has grammars for over 40 languages. Both systems have been found useful in applications that need to process several languages. Their core idea is to exploit levels of abstraction. On a universal level, the grammar is "the same" for different languages. On a more concrete level, the details that distinguish languages are expressed accurately. In spite of looking totally different superficially, UD and GF turn out to be very similar in their choice of universal concepts, and they can be combined in different useful ways.