

Abstract:

Principles of Processing and the Distribution of Complexity in Natural Language

In this talk, I will present new ideas and preliminary results of my research on human language processing. The main goal of my work is to formally describe the human capacity of linguistic processing, or more general, abstract symbol processing, under the assumption that it is systematic, and with the goal to account as accurately as possible for the richness (as well as restrictions) of symbol systems as natural language, while having to assume just a minimal amount of abstract symbol manipulation.

Technically, I implement this idea in a restrictive algorithm, i.e. an algorithm which uses a given grammar formalism (in our case, Combinatory Categorical Grammar), but analyzes only a sublanguage of the language the grammar generates. The guiding principle for the design of the algorithm is linearity: it should be able to analyze any string in a number of steps growing linearly in terms of the length of the string, or else reject it. This is achieved with a chunk parser which works incrementally on two levels: where dependencies between two adjacent elements cannot be handled locally on the string level, the first element is transferred to the chunk level; when the dependency can neither be handled on the chunk level, the string cannot be properly analyzed. The main constraint is that on either level, we always have to combine elements in order to yield a single function (by means of function application and our combinators). The class of languages which can be analyzed shows some interesting properties: on a local level, we have the full power of a mildly context-sensitive grammar formalism; on a more global level, we can analyze a class of languages which properly contained in the class of context-free languages, and properly contains the class of regular languages.

This property, among others, makes strong and plausible predictions for human sentence processing, and allows to account for plenty of observations made by linguists which normally are quite challenging to formal linguistics; they cover acceptable ungrammaticality (as in resumption and functional uncertainty in long-distance dependencies), as well as "unacceptable grammaticality" (as in multiple center embedding).

As an outlook, I will present some evidence how chunking affects the interaction between modules of the grammar, and facilitates a description of the interaction of linguistic subsystems which does not depend on the (at least in spoken language pretty arbitrary) unit of the sentence.