

PPP Project meeting summary Chemnitz, April 2005

The first project meeting aimed at providing a common ground for discussion on the issue of complexity in linguistics/complexity in language. A second topic was an evaluation of current research projects of the individual project members regarding their applicability to a notion of complexity. Thirdly, we agreed on a schedule for further meetings.

The attempts at defining complexity does not always provide a plausible basis for linguistic analysis. Problematic are especially ad hoc definitions or household items that lead to folk views of complexity without powerful constraints. Problematic are further abstract definitions quite removed from the practice of linguistic description mechanisms. We tried to define a smallest common denominator of complexity with a definition from natural sciences/information science. Here, complexity is a measure of randomness within structure. Structure is defined as the presence or absence of information. Any structural phenomenon can be described and enumerated using algorithms (example: physics, gravity etc. governed by physical laws that can be captured in relatively simple equations). Therefore, the less algorithms are needed to fully enumerate a phenomenon the less complex it is. More complex phenomena are described by a larger number of algorithms. A phenomenon counts as structurally random when the number of algorithms to enumerate e.g. a sequence of informationally diverse items is equal to the number of items themselves. Structures of this complexity are found in nature (number theory) but not in artefacts. Artefacts like natural language are filtered and therefore determined by the (considerable) complexity of human language competence (speak: brain) which is complex but not infinitely complex.

Trying to match this definition with linguistic description turns out difficult only at first glance. Any model in linguistics accounts for a certain spectrum of phenomena, simple clauses e.g. receive less grammatical effort. It therefore makes sense to subdivide the very general notion of complexity into

- 1) structural complexity
- 2) processing complexity/cognitive complexity
- 3) "everything-else" complexity

- 1) as for structural complexity, it has been suggested to find patterns in syntactic occurrences based solely on the tags of corpus texts. As a principal method of study this computes a gradient of recurring patterns. The larger the chunks of recursion, the less complex is the text. This could be related to texts that look complex on the surface but rather are not, cf. legal texts. The example of relative clauses brought the discussion to the point where structural complexity alone can not account for linguistic complexity alone. This also raised comments and discomfort with semantic/pragmatic approaches and clearly shows a desire to include interpretation and "aboutness" of linguistic symbols into consideration of complexity.
- 2) Processing complexity has to do with the perception and conceptualization of structure. The example of recursively embedded sentences shows that processing breaks down for algorithmically very simple structures very easily. Concepts also govern e.g. choice of relative pronouns and interference effects account for differences in complexity. Perceived complexity of first vs. second language was briefly discussed and rejected as a means of research as it does not correlate with the order of acquisition.
- 3) "Everything-else" complexity has strong pragmatic undertones and needs further specification in order to account also for structures beyond the clause level. Text colonies and retrieval of items from a Functional Sentence Perspective point of view were mentioned here.

Participants agreed on pursuing different strands of research that cut across linguistic subdisciplines and will be reconciled by a common notion of complexity which will resemble the one mentioned above but will be subject to further specification. The following subprojects therefore tie in with the general project topic:

- a) relative clauses as embedding/subordinate structure are a rich testing ground when considered in a contrastive English-German vs. English-Czech way.
- b) Causation is grammaticalized differently in different languages. It would therefore be interesting to look at texts of learner English and compare their strategies of expressing causation with special regard to lexical causation
- c) Response elicitors in spoken English are distributed according to non-trivial underlying rules that to uncover will shed some light on the complexity of spoken vs. written language
- d) The interrelationships between structurally similar components as supplementary clauses in text colonies (such as UNESCO resolutions) that determine complexities on a higher intertextual order
- e) "Linkers" in academic texts can be classified on different levels of complexity, e.g. concessives where defined as expressing the "most complex" semantic relationship (Kortmann, 1991)
- f) Conceptualization of English present perfect by second language learners underlies different parameters the interaction of which is nontrivial to uncover

The next major project meeting in Brno will summarize the attempts sketched here. We will try to find different resources of data for the resulting subprojects of project members (as: recordings of Germans/Czechs speaking English; narrative texts from German/Czech composition classes; concept judgments on present perfect; UNESCO texts etc.). Some tentative overview of texts with the means of corpus analyses can show that complex semantic items (as causation across longer distances in a text using all properties of textuality etc.) will display fewer occurrences. The same items should also exhibit longer processing times in elicitation tests. Other examples are the complex of stance/perspective/point of view phenomena all expressing commitment of a speaker toward a particular state of affairs.