Parsing Yoruba serial verb constructions: an attempt in the LFG framework and its testing on a large corpus of Yoruba

Nicolas Aubry

CNRS - France

While creating a Lexical Functional Grammar (LFG) representation of Yoruba grammar, one of the most difficult tasks facing the linguist is the setting up of the rules concerned with the analysis of serial verb constructions (SVC) – of course this would be true for any other syntactic formalism and serializing language as well. Typically, by dramatically broadening the potential rewritings of the VP (e.g. VP → VP VP) constituent at the c-structure level, one increases the risk of overgeneration and, in the case of an analysis by an automatic parser (which is the case of this study), of giving far too much (erroneous) interpretations to a single sentence. One is hence left with the mandatory task of constraining those c-structure rules.

Given the absence of a priori lexical restrictions preventing any verb of occuring in a serializing clause (which strongly distinguish them from the most common subordinate structures, e.g. those triggered by a « control » verb), a complex multilevel (involving c-structure, f-structure and the lexical entries of both verbs and nouns) set of interweaving constraints must have been set up. I argue for the suitability of LFG to formalize this kind of complex constraining, permitting notably to correctly assign the grammatical relations (which is a central issue in the analysis of serialization) in an « elegant » way. The LFG grammar thereby created having been tested on an original 500000 words diachronic corpus consisting of a century of Yoruba press, I then discuss the results of this test concerning serializing clauses.