GeLexi Project: Sentence Parsing Based on a GEnerative LEXIcon

Gábor Alberti, Judit Kleiber, and Anita Viszket

University of Pécs, Faculty of Humanities, Linguistics Department
H7624 Pécs, Ifjúság útja 6., Hungary
gelexi@btk.pte.hu
http://lingua.btk.pte.hu/gelexi.asp

Keywords: parsing of (Hungarian and English) sentences, Prolog, lexicalist generative grammar, DRS (discourse representation structure)

The principal aim of our Pécs research team, called *GeLexi*, is to verify that computational linguistics is worth returning from the nowadays wide-spread attitude characterized by "shallow parsing" (which is held to save expenses) to the pure theoretical (generative) linguistic basis [15, 21].

Our crucial argument relies on a double (parallel computational and linguistic) chance: to use simultaneously, on one hand, a significantly greater number of huge patterns than earlier due to the immense increase in memory capacity [22], and to work out a formal grammar, on the other hand, showing the distribution of capacity advantageous in modern computer science (in harmony with the development mentioned above): "minimal processing – maximal database". This latter chance has something to do with the sweeping *lexicalist* turn [13, 14, 16, 18, 19] in generative linguistics, which used to be chiefly "process-oriented" (i.e. syntax-centered) in its first period; the current attitude can be characterized by two mottoes of Joshi's [18], the father of *mildly context-sensitive* grammars [21]: "Complicate Locally, Simplify Globally", and "Grammar ≈ Lexicon".

What we propose is a new sort of generative grammar, *GASG* ("Generative/Generalized Argument Structure Grammar", defined in [6] and demonstrated in a wide range of papers [1-11]), which is more radically "lexicalist" [19] than any earlier one. It is a modified Unification Categorial Grammar [19, 17], from which even the principal syntactic "weapon" of CGs, Function Application, has been omitted. What has remained is *lexical sign* and the mere technique of *unification* as the engine of combining signs.

Our GASG-parser, in accordance with the basic task of every generative grammar [15, 21], decides whether a sentence is *grammatical*, and then provides a *morpho-phonological* analysis (based on a "Totally Lexicalist (approach to) Morphology" launched in [5]), a compilation of *grammatical relations*, and two kinds of *semantic* representations: a DRS [17] completed with information about its embedding in interpreters' information state also formulated as a DRS [2], and a network of *copredications*, useful in translation [7, 8].