

# Why Implementors of Practical NLP Systems Can not Wait for Linguistic Theories

## Remarks and Theses

Dietmar Roesner  
c/o Project GENESIS  
GMD F4, Dolivostr. 15  
D 6100 Darmstadt  
West Germany

A) There is no doubt: the idea of taking an accepted linguistic theory (lt) as the basis of the implementation of a practical natural language processing (nlp) system is very attractive.

But: In reality there are no linguistic theories available that may be simply "taken from the shelf" and applied in such a way. Some reasons for this:

- There is often a mismatch between the problems dealt with extensively in a given linguistic theory and their frequency of occurrence in real data.

- On the other hand the chosen linguistic theory or grammar may tell you only little or nothing about phenomena that are predominant in your application (e.g. large nominal groups in technical texts).

- Practical systems need a substantial if not complete coverage of the syntax of the nl they are built for. In general the language fragments covered by linguistic theories are not sufficient.

- Even if you are able to start with structures in accordance with a linguistic theory when building a practical nlp system you will have to augment the formalism: e.g. as soon as you have to process real data and not only text book cases a lot of conventional material has to be taken into account as well: format of dates, measurements, etc.

- Linguistic theories very often have been extensively applied to problems taken from a specific natural language (in most cases: English). To transfer those theories to other languages with different problems may be hard (if not impossible).

B) Another difference between theoretical and engineering problems lies in the fact that building a practical nlp system always forces you to take a lot of additional constraints into account that are negligible from a theorist's viewpoint:

- For the implementation, usage and maintenance of practical nlp systems dedicated software tools are needed e.g. for editing, debugging and update of grammars and lexica.

- Practical nlp systems are often evaluated by users (and referees) that are non-linguists. According to our experience those people tend to take "superficial" aspects that have little or no relevance for linguists as a basis for their judgements; e.g. if or if not a system for German is able to handle umlauts correctly or how you may type in to a system for Japanese.

- Practical nlp systems inevitably need a substantial lexicon and - due to the generativeness of nl with respect to vocabulary - additional techniques for handling unknown lexical items.

C) Some of the above holds especially for work in nl generation:

- Linguistic theories are primarily concerned with aspects of (syntactic) analysis.

- Some of the hard problems for analysis (e.g. PP attachment) are not equally problematic in generation.

- In general nl generation is an interdisciplinary task involving a lot of non-linguistic decisions.

D) Theorists tend to restrict their approaches to the very techniques available within their theories. In practical nlp systems it may be fruitful to freely combine elements from distinct "linguistic schools". The morpho-syntactic front end generator FREGE [Emele 87] is a case in point: FREGE takes functional grammatical structures - comparable to those from LFG [Kaplan, Bresnan 82] - as input and working structures but provides means to specify constituent ordering as linear precedence relations similar to GPSG [Gazdar et al. 82].

E) What about the following analogy: The relation between theoretical linguistics and "language engineering" should parallel that between mathematics and (civil) engineering.

F) Linguistic theory might profit from problems attacked in practical systems as well because (as with mathematics and problems in engineering) these problems may give motivation for future research. Practical work in multilingual generation - i. e. attempts to generate different languages from the same semantic structures - may for example give a new perspective for work in contrastive linguistics.

To sum up: One should not wait for theoretical solutions since theory might not attack some problems until they are of practical relevance.

## References

[Emele 87] Emele, M. (1987): "FREGE - Ein objektorientierter Front-End-Generator", in: K. Morik (Ed.) "GWAI-87 - 11th German Workshop on Artificial Intelligence", Proceedings, Springer-Verlag

[Gazdar et al. 85] Gazdar, G., E. Klein, G. Pullum & I. Sag (1985): "Generalized Phrase Structure Grammar", Cambridge: Harvard University Press

[Kaplan, Bresnan 82] Kaplan, R. & J. Bresnan (1982): "Lexical Functional Grammar: A Formal System for Grammatical Representation", in: Bresnan, J. (ed.) (1982): "The Mental Representation of Grammatical Relations", Cambridge (Mass.): MIT Press