

Linguistic Bases For Machine Translation 1)

Christian Rohrer

Institut für Linguistik
Universität Stuttgart
Keplerstraße 17
7000 Stuttgart 1

My aim in organizing this panel is to stimulate the discussion between researchers working on MT and linguists interested in formal syntax and semantics. I am convinced that a closer cooperation will be fruitful for both sides. I will be talking about experimental MT or MT as a research project and not as a development project.[1]

A. The relation between MT and theoretical linguistics

Researchers in MT do not work with linguistic theories which are 'en vogue' today. The two special issues on MT of the journal *Computational Linguistics* (CL 1985) contain eight contributions of the leading teams. In the bibliography of these articles you don't find names like Chomsky, Montague, Bresnan, Gazdar, Kamp, Barwise, Perry etc.[2] Syntactic theories like GB, GPSG, LFG are not mentioned (with one exception: R. Johnson et al. (1985 p.165) praise LFG for its 'perspicuous notation', but do not (or not yet) incorporate ideas from LFG into their theory of MT). There are no references whatsoever to recent semantic theories.

On the linguistic side one notices a similar lack of interest in MT or in the theory of translation in general. In the latest books on GB, LFG or GPSG one will look in vain for any references to the theory of translation. Or, to give another example, look at the research goals of the recently created Center for the Study of Language and Information (CSLI) at Stanford. The aim is to develop "integrated theories of language, information and computation". (CSLI Report No. 84-16, p.2). At CSLI philosophers, linguists, computer scientists and psychologists collaborate intensely on numerous projects, however MT is not among them. There is only one notable exception (M. Kay) which confirms the rule. None of the other 40 staff members has ever been involved in an MT project.

This does not mean that there is no research in translation going on at all, quite on the contrary. A growing number of countries are creating chairs in the theory of translation, there are symposia on the theory and practice of translation (L. Grähls et al. 1978), and there exist innumerable papers on practical problems of translation, which are written primarily for interpreters and teachers of a second language. The efforts in this area can be characterized by key words like contrastive linguistics and literary theory. Within contrastive linguistics the emphasis lies on lexicology. (Typical examples which are discussed are pairs like Engl. *to know*, Fr. *savoir*, *connaître*, Ital. *scala*, Germ. *Leiter*, *Treppe*.) Yet, there is little detailed formal work on contrastive syntax. (For an exception see Hawkins (1986)). The results in contrastive linguistics in general and in contrastive lexicology in particular will certainly be of great value for writing transfer dictionaries and specific transfer rules. The research undertaken in this area, however, has not produced sufficient results to build a coherent formal theory of translation on. Finally there is the philosophical tradition, best illustrated by Quine's famous 'Meaning and Translation' (1959) and Montague's 'Universal Grammar' (1970). In this context translation means translating from a natural language into a logical language. This sort of translation has had an enormous influence on semantic theories used in linguistics but again there are very few connections with MT (for an exception see Landsbergen (1982)).

B. Why is there so little interaction between MT and theoretical linguistics?

- (1) Most MT projects were set up as development projects. There was just not enough time or money to do basic research.
- (2) MT projects were often directed by computer scientists who underestimated the complexities of natural language.
- (3) Analysis grammars for MT were often written in a formalism which made communication with the non-initiated impossible.
- (4) Theoretical linguists thought and still think that MT is too difficult, that high quality MT is impossible. Some argue on the basis of syntax: if we don't even have a satisfactory syntax for comparatives in English, how can we translate English comparatives into German. Some argue on the basis of semantics: satisfactory semantic analyses exist only for very small fragments of natural languages. How are we to translate German tense forms into French tense forms, if nobody has yet presented an adequate semantic description for French tenses.
- (5) Linguists seem to have a one-sided (lop-sided) perspective, working nearly exclusively on analysis. In LFG one parses sentences and builds up functional structures, in GPSG sentences are parsed and translated into formulas of intensional logic, hardly anyone knows how to generate from f-structures or from logical formulas.
- (6) If there is little research on generation, there is even less on *transfer*. Under the influence of Chomsky and (to a smaller extent) Montague linguists started the search for Universal Grammar. Transfer, however, presupposes the comparative study of two (or more) languages. Linguists working with recent syntactic theories don't seem to be interested in questions like:[3] Is language A syntactically more ambiguous than language B? If a string S is ambiguous in language A, is there a string S' in language B which presents the same number of ambiguities? Is it really easier to translate between two languages which are closely related (say Italian and French) than between distant languages (say Japanese and German)? Under what conditions is it possible to translate on the basis of syntactic structure, and under what conditions do we have to use semantically based transfer rules?

C. What can we do to stimulate interaction between MT and theoretical linguistics?

I would like to argue that MT should be based on a linguistic theory which can be expressed in a formalism whose formal properties are well understood, in other words, a formalism for which there exist results on decidability, generative capacity and recognition complexity. The linguistic theory itself, independent of its formalisation, should be well accepted within the linguistic community. Possible candidates for such theories are GB, GPSG, LFG etc.[4] What are the advantages of working with such a theory?

(1) Advantages for the linguist

The MT linguists can use directly analyses which have already been proposed for the phenomena he is trying to describe. If a linguist writes a GPSG grammar for a fragment of English and wants to include coordination, he can just use the analysis proposed by I. Sag et al. (1984).

He can discuss the insights expressed in his own analyses with linguists outside the MT groups.

Since the formalisms which are used for expressing linguistic theories are closely related (Shieber (1985) has even shown that the formalisms, in which the different UGs are expressed, are irreducible to a large extent), a particular analysis can be transferred from one theory into the other. The treatment of long-distance dependencies (*wh* - movement) in LFG for instance looks formally very similar to that in GPSG.

Linguists working outside of MT could use the implementation of LFG, GPSG etc. as grammar testers.

Theories like LFG, GPSG offer a well-defined interface to semantics. To name just one particular problem we are working on at the moment, it thus becomes possible to test the adequacy of an analysis of the French tenses and of the English tenses by writing transfer rules which map a semantic structure, which represents the temporal information contained in a small French text, into the corresponding semantic structure for English.

In UGs the lexicon plays a central role. A lot of linguistic information is located in the dictionary entries. The rich lexical information collected by contrastive linguists could be incorporated into a transfer lexicon written according to the principles of UGs.

(2) Advantages for the computer scientist and linguist

If linguists can write their grammars in a formalism whose mathematical properties are well understood, then the programmer will have fewer problems implementing the grammar. Results on formal properties of grammar formalisms can guide the programmer in his search for appropriate algorithms. Furthermore, in the same way in which a linguist working within LFG can benefit from the intuitions expressed in linguistic analyses within GPSG or GB, a programmer implementing LFG can benefit from insights contained in implementations of related formalisms. It is therefore not surprising that F. Pereira (1981), W. Frey and U. Reyle (1983), L. Karttunen (1985) and J. Calder et al. (1986) all use "gap threading" for implementing long-distance dependencies. The authors are working with different theories (Extrapolation Grammars, LFG, Helsinki Unification Grammar (HUG) and Unification Categorical Grammar (UCG)) and different programming languages (Prolog and Lisp) and nevertheless the algorithm they chose is the same. Let me give you a further example. In the Prolog implementation of LFG by Dörre and Eisele (this volume) there are some problems with negative and disjunctive constraints. Such constraints are linguistically well motivated but difficult to implement. Now if anybody within the UG community comes up with a good implementation of negative and/or disjunctive feature specification, then this new implementation can be incorporated.

Results on recognition complexity may help to locate the source of the complexity and suggest ways to reduce complexity either by modifying the linguistic theory or by showing that the "worst case" is linguistically (empirically) irrelevant (in other words, the "worst case" will never arise in practice).

A famous example, where the linguistic theory was changed after a proof of its generative power had been found, is Peters and Ritchie's work (1973) on the generative power of TG and the subsequent change of the theory (constraints on deletion rules).

For GPSG Peters and Uszkoreit (1983) have examined the role of metavarules and essential variables (variables which range over all strings of terminal and nonterminal symbols). They proved that unconstrained Metarule Phrase Structure Grammars have Turing machine power. This result has led to intense research for constraints that are linguistically motivated and descriptively adequate.

For LFG there is a conjectured result by Berwick: certain Lexical-Functional Grammars can generate languages whose recognition time is very likely computationally intractable, at least according to our current understanding of algorithmic complexity." (Berwick 1982, p.98)

Conclusion: Basically any argument in favor of a declarative, simple, expressive, flexible linguistic formalism with a rigorous mathematical foundation can be adduced here.

D. Possible objections against using Unification Grammars or other recent linguistic theories for MT

Are there linguistic phenomena, which make it impossible to use UGs for parsing and generation? Someone might argue that in MT there exist specific problems, which have never been taken into consideration by theoretical linguists, and which cannot be handled with UGs. Personally I hope that whatever additional problems may arise can be handled in the transfer component. If UGs provide us with the tools to write adequate grammars for individual languages, why should it not be possible to use these grammars for MT?

R. Johnson et al. consider the possibility of assuming a current linguistic theory and implementing it directly for EUROTRA. They reject this solution "because there is not sufficient practical evidence of a single theory that encompasses translations between all pairs of the Community languages" (1985, p.165). This rejection can be interpreted at least in two ways: (a) there is no linguistic theory in which all the languages of the Community have been described, (b) recent linguistic theories have not been used for (automatic) translation between all pairs of the Community languages. Of course, both interpretations are correct; however, on the basis of this argumentation one could reject any existing linguistic theory as insufficient for MT. This reasoning, however, would force us also to reject any new linguistic theory because there would be even less "practical evidence".

So far only fairly small grammars have been written and implemented in the UG formalism. What will happen if these toy grammars develop into big grammars with very large dictionaries? Won't we need special control mechanisms? Since the UG formalism is well understood, I can't see why one should have more problems if the grammars get bigger than with any other formalism. We hope that the contrary will be the case, namely that there will be fewer problems.

There is still another side to the problem of "toy grammars". Theoretical linguists have been working with a fairly narrow set of data. An ambitious young linguist today is more likely to propose the 127th version of the passive construction than to write a detailed grammar for complex NPs like "die Einfuhr von Jute aus Ländern der dritten Welt zum Ausgleich des Nord-Süd-Gefälles". Will theoretical linguists ever sit down and actually write big grammars? Why has there been no grammar writing project since the attempt by R. Stockwell et al. (1968) to collect all TG analyses into one grammar?

UGs have been used for analysis, but there are practically no results for generation. This lack of experience in generation is a serious drawback for the whole field of computational linguistics. Fortunately the situation is gradually changing. The organizers of the International Workshop on Language Generation (Stanford 1984) write: "More papers in this subfield [generation] have been published in the last four years than in its entire previous history."

In MT generation has played a minor role. In a system like SUSY (Maas 1985) for instance there is no explicit grammar for the generation of German. The Japanese Government Project for MT has a well developed generation component (Nagao M. et al. (1985)), but it is difficult for the linguist to find out how this generation grammar would look in a declarative format. So, even if there are no results on generation with UGs we are not much worse off than anybody else working with another linguistic theory.

The most serious objection I see concerns the transfer component. I have been advocating the use of linguistic theories which can be expressed in well-defined formalisms, but so far there doesn't seem to exist a theory of transfer for UGs. Well this is not entirely correct. M.Kay claims that his Functional Unification Grammar (FUG) "provides an opportunity to encompass within one formalism and computational system the parts of machine translation systems that have usually been treated separately, notably analysis, transfer and synthesis." (Kay, 1984, p.75).

E. Conclusion

MT as a research project cuts across traditional disciplinary and institutional boundaries. Theoretical linguists, lexicographers and computer scientists must cooperate in order to solve some of the problems of (machine) translation. Grammars and dictionaries used in MT should be based on linguistic theories which are widely accepted in the linguistic community. The linguistic theories should be expressed in formalisms that fulfill the requirements of a good programming language. An open and transparent discussion between linguists and computer scientists will lead to a realistic assessment of the possibilities and prospects for MT. We should avoid a reinterpretation of the motto 'traduttore traditore'.

Footnotes

- 1 Support for work on this paper was provided by a grant from the BMFT.
- 2 The choice of names and of theories is for exemplificatory purposes. We do not try to be exhaustive in any sense. No evaluation is implied by the presence or absence of any particular name, theory or publication.
- 3 Fortunately the situation is changing now. Since Chomsky's Pisa Lectures there is a growing interest in comparative syntax. Maybe some day a theory of translation will emerge from the theoretical and practical insights from these comparative studies within the framework of GB. It is too early to construct today a theory of translation on the basis of GB, unless this theory made reference to GB only in a metaphorical sense like E. Nida did in his proposal to use the Aspects model for translation and to do transfer on "deep structures" (Nida (1969)).
- 4 In the following I will use the term Unification Grammar (UG) as hyperonym for GPSG, LFG, FUG, HPSG etc., i.e. for grammars which have a contextfree skeleton and an operation of unification. In grammars of this type, syntactic rules and lexical entries can be expressed as sets of attribute-value pairs. The value of an attribute can again be a set of attributes and values. Later on I will argue for the use of UGs for MT. This does not mean that GB or Joshi's Tree Adjoining Grammars could or should not be used for experimental MT. I just happen to have a personal preference for UGs.

References

- Berwick, R.C. (1982), 'Computational Complexity and Lexical Functional Grammar', *AJCL* 8, 3-4, pp.97-109.
- Calder, J. et al. (1986), *Problems of Dialogue Parsing*, MS., Center for Cognitive Science, Edinburgh.
- Chomsky, N. (1981), *Lectures on Government and Binding: the Pisa Lectures*, Dordrecht.
- Frey, W., Reyle, U. (1989), 'A Prolog Implementation of Lexical Functional Grammar as a Base for a Natural Language Processing System', *Proceedings of the First Meeting of the Association for Computational Linguistics, Pisa*, pp.52-57.
- Gazdar, J. et al. (1985), *Generalized Phrase Structure Grammar*, Oxford.
- Gribs, L. et al. (ed.) (1978), *Theory and Practice of Translation*, Nobel Symposium 30, Stockholm.
- Hawkins, J. (1986), *A Comparative Typology of English and German Unifying the Contrasts*, London, Sidney.
- Johnson, R. et al. (1985), 'EUROTRA: A Multilingual System under Development', *CL11*, pp.155-169.
- Joshi, A.K., Levy, L.S., Takahashi, M. (1975), 'Tree Adjunct Grammars', *J. Comp. and Sys. Sc.* 10, pp.136-163.
- Kaplan, R., Bresnan, J. (1982), 'Lexical-Functional Grammar: A Formal System for Grammatical Representation', in: Bresnan, J. (ed.), *The Mental Representation of Grammatical Relations*, Cambridge Mass., pp.173-281.
- Karttunen, L. (1985), *IUG: a Development Environment for UGs*, MS., Stanford
- Kay, M. (1984), 'Functional Unification Grammar: A Formalism for Machine Translation', *Coling 84*, pp.75-78.
- Landsbergen, J. (1982), 'Machine Translation Based on Logically Isomorphic Montague Grammars', *Coling 82*, pp.175-181.
- Maas, H. (1985), 'Zur Entwicklung von SUSY-II', in: U. Klenk (ed.) *Kontextfreie Syntaxen und verwandte Systeme*, Linguistische Arbeiten, Tübingen, pp.107-117.
- Montague, R. (1970), 'Universal Grammar', *Theoria* 36, pp.373-398.
- Nagao, M. et al. (1985), 'The Japanese Government Project for MT', *CL 11*, pp.91-110.
- Nida, E. (1969), 'Science of Translation', *Language* 45,3, pp.483-498.
- Peters, P.S., Ritchie, R.W. (1973), 'On the Generative Power of Transformational Grammars', *Inf. Sc.* 6, pp.49-83.
- Pereira, F. (1981), 'Extrapolation Grammar', *CL 7*, pp.243-256.
- Quine, W. (1959), 'Meaning and Translation', in: Brower, R. (ed.), *On Translation*, Cambridge Mass., pp.148-172.
- Shieber, S. (1985), 'Separating Linguistic Analyses from Linguistic Theories', Ms., Stanford.
- Stockwell, R. et al. (1968), *Integration of Transformational Theories on English Syntax*, Los Angeles.
- Uszkoreit, H., Peters, P.S. (1983), *Essential Variables in Metarules*, *SRI Tech. Note* 305.