NEW DIRECTIONS IN MT SYSTEMS

Christian Rohrer

Institute for Natural Language Processing, University Stuttgart, FRG

Progress in hard- and software as well as new results in theoretical linguistics constitute optimal conditions for the development of practical MT systems. We want to enumerate briefly some new directions of research for MT systems of the future.

Declarative Grammars

Recent developments in grammar formalisms allow the linguist to write grammars independently from the implementation. Linguistic descriptions for MT systems or other applications in natural language processing can now be written in close interaction with theoretical linguists. The results are easier to understand and can be modified or expanded by researchers who are not involved in the implementation.

Reversible Grammars

Declarative grammars, especially within the paradigm of unification formalisms, are now being used in experimental systems for parsing and generation. The same grammar can be used to derive an abstract representation from a string (analysis) and to derive a string from an abstract representation (generation). In other words, the grammar is reversible. There are already some very promising results.

Reversible MT Systems

If grammars are reversible, why not make the whole MT system reversible? This implies that transfer is reversible or that an interlingua is used. First attempts to make transfer reversible are under way. In such systems, transfer relates two language-dependent semantic structures. So far, however, all the relevant knowledge is contained in the transfer lexicon.

Multilevel Transfer

It has been shown that phonological, morphological, syntactic and contextual aspects contribute to the meaning of a sentence. Therefore, in MT too, one must provide the possibility to refer simultaneously to different levels. These considerations have led to the concept of multi-level transfer, a concept which is now being implemented by several research groups around the world.

Multilingual MT Systems

MT systems for organizations such as the European Community must be able to translate between many different languages. MT based on directed transfer between n languages requires n * (n-1) different transfer components. Intensive research in contrastive linguistics is necessary in order to exploit the syntactic and semantic parameters common among languages of the same family (like the Romance languages, or between Japanese and Korean). Furthermore, work on formal theories of grammar has considerably sharpened understanding of linguistic universals - principles that are common to all languages. In the future, results from 'Universal Grammar' (Chomsky) may be used in MT. Some prototype systems inspired by this approach are already being implemented.

Semantic representations, which are more abstract than the syntax-oriented representations of earlier MT systems, will also contribute to the construction of multilingual systems by reducing the tasks of the transfer component.

Knowledge Representation and Disambiguation

Resolution of lexical and syntactic ambiguities is a major problem of MT. It is well known that many ambiguities cannot be resolved without common-sense knowledge. Consider for instance an example from the German Federal Railway Table: "mothers with children under four". Only our world-knowledge tells us who is four years old.

Now, independently of MT, great efforts are made by AI researchers to represent common-sense knowledge for expert systems, computer aided instruction, and knowledge processing technology in general. It can be expected that advances in knowledge representation will contribute significantly to the resolution of ambiguities. Knowledge representation will be especially effective if the MT system is restricted to a special domain of application. It

should not be forgotten that there is a vast amount of work in formal semantics which is also relevant in this respect.

Automatic Translation of Spoken Language

Spoken language has priority over written language historically, functionally and biologically. Even today in the most literate of modern industrialized and bureaucratic societies, the spoken language is used for a wider range of purposes than the written. Therefore it is critical, in processing natural language, to be able to handle spoken language. The largest attempt to combine speech-analysis and -synthesis with MT is presently under way at ATR in Japan. The research results available so far indicate that interactive automatic translation of spoken language is possible in a limited domain of application and that robust grammars can be implemented successfully.

Technical writing and controlled language

The better a text is written the easier it is to translate. This is true for human and machine translation. We have to teach technical writing at university level and we have to investigate the principles for writing texts which are easy to understand for humans and computers. This investigation should be carried out in parallel with research in 'controlled' language, i. e. in fragments of natural languages, where all syntactic and semantic constructions are well understood and described explicitly, but which are nevertheless rich and expressive enough to permit the formulation of good texts.