

THE LRC MACHINE TRANSLATION SYSTEM: AN OVERVIEW OF THE
LINGUISTIC COMPONENT OF METAL

Winfield S. Bennett

Linguistics Research Center, The University of Texas at
Austin

Although there has been work on machine translation since the early 1950's, it is only in the last few years that MT systems have become capable of producing output anywhere near the quality needed for wide-spread use. The Linguistics Research Center (LRC) at the University of Texas at Austin has developed a working machine translation system. METAL, the actual translation system, produces quality English translations for German technical manuals. Experiments conducted with up to 200 contiguous pages of text have been most encouraging; the LRC plans to have a production system by early 1983.

METAL is a modular system in which the lexicons, grammar, and analyzer are distinct packages, each of which may be modified without affecting the other two. This paper will present an overview of METAL's linguistic component, giving a general overview of the scope and format of the METAL lexicons and a somewhat more detailed description of the METAL grammar. It will then present a description of the translation process using a simple German sentence, and conclude with a discussion of the current state of the linguistic component, along with comments on what improvements are necessary in the further development of the system.

METAL Lexicons

This section will briefly describe the scope and format of the system lexicons. METAL employs monolingual lexicons for both the source and target languages, and bilingual lexicons to equate each source language word with a corresponding target language equivalent. The bilingual or "transfer" lexicons permit a word in the source language to be translated into any of variety of words in the target language depending on nuances of grammatical context. The lexicons are feature-based, consisting of a series of syntactic and semantic features and values for each lexical entry. The paper will discuss this feature-based system in general, looking at the format of both the monolingual and bilingual ("transfer") lexicons. This section will conclude with comments on the coding procedures developed, but not yet fully implemented, for use in METAL lexicons.

METAL Grammar

This section will describe the format for the grammar rules used in the analysis of source language texts and the synthesis of corresponding target language strings. METAL grammar rules are augmented context-free phrase structure rules with a wide range of options allowed in the augmentation parts.

The METAL approach to grammar is rather different from the "standard" approach found in many recent articles. In the recent work by linguists such as Chomsky, Bresnan, and Gazdar, the main thrust has been to establish a theoretical framework which may then be applied to human language; METAL's approach has been the opposite, i.e., to seek solutions to real language problems and to use these solutions in the development of a theoretical framework.

The grammatical framework allows the linguist to devise tools when they seem appropriate, and to discard them if they turn out to be unnecessary. In this section particular emphasis will be given to the various augmentation parts of the grammar rules, particularly how these parts are used to constrain rules, to build syntactic trees, and to generate the appropriate word order for the target language. This section will conclude with a discussion of METAL transformations and of the case frames mechanism, which is used to disambiguate strings.

Sample Translation

To illustrate the operation of the linguistic component within METAL, the paper will then examine some significant steps involved in the analysis and synthesis of a German sentence. This examination will point up some of the typical problems in German - English machine translation, and the solutions used in METAL. Using representative lexical entries and grammar rules to demonstrate the process, the paper will illustrate the production of the correct English translation for the sample sentence.

Conclusion

METAL is capable of producing quality machine translation of a large number of German sentence into English, but the system needs more development before a production version is available. In this section the paper will describe the limitations on and problems with the METAL linguistic component in its present form, and will outline the improvements currently planned to overcome these limitations and problems. Particularly important in this regard is the need to develop the semantics of METAL from its present state into a full scale MT semantics.