SESSION 8A: MACHINE TRANSLATION

Jerry R. Hobbs, Chair

Artificial Intelligence Center
SRI International
Menlo Park, California 94025

The three papers in this session exemplify three different approaches to machine translation. Each addresses a significant problem encountered within the approach.

In the dark ages of machine translation, it is said that researchers attempted a Direct Approach—word-for-word substitution. But this could not have survived much beyond the first inspection of the output. The most common approach adopted in working machine translation systems is the Transfer Approach illustrated in Figure 1. A text in a source language is analyzed to some depth, producing, for example, a parse tree or a logical form for the sentences in the text. Then transfer rules are applied to this representation to produce a representation at the corresponding level for the target language. The text is then generated in the target language.

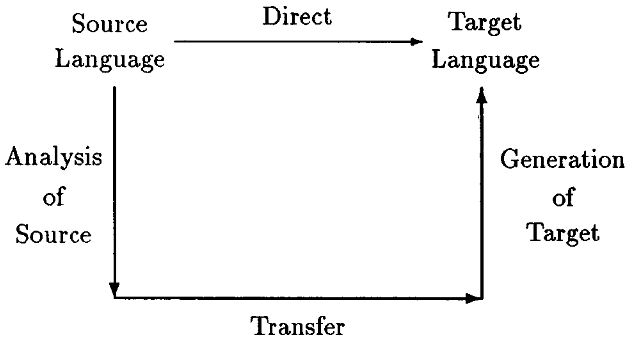

Figure 1: Transfer Approach to Machine Translation.

A French-to-English translation system might have a transfer rule like the following. The French phrase

la rémunération du temps supplémentaire

has the following syntactic structure:

$[_{NP}$ la rémunération $[_{PP}$ du
$[_{NP}$ temps supplémentaire]]]

The corresponding English phrase

overtime pay

has the syntactic structure

$[_{NP}[_{N}$overtime] pay]

The transfer rule would specify how fragments of the French parse tree mapped into the corresponding fragment of the English parse tree, for example, how the PP in French maps into the prenominal noun in English. The rule would be stated with whatever lexical generality is appropriate.

There are a number of classical problems with the Transfer Approach, principally arising when the two languages express the same concept in very different ways syntactically. For example, what is expressed by the main verb in one language may be expressed adverbially in another. Conjunction reduction may be possible in one language, while lexical factors make it impossible in another. The paper by Kinoshita et al. addresses many of these problems and describes how they can be handled within the Transfer Approach.

Another approach, long advocated but rarely given extensive implementation, is the Interlingua Approach, illustrated in Figure 2. Here one does a much deeper analysis on the source-language text, to the level of a language-independent conceptual representation called Interlingua. There are two advantages to this approach. First, a text must often be analyzed to a conceptual level in any case to achieve an adequate translation. The text must be understood before it can be translated. Second, when there are many languages one must translate among, we need only define the mapping between each language and the Interlingua, rather than specifying the transfer rules for every pair of languages.

One problem with the Interlingua Approach is the difficulty, if not impossibility, of devising an adequate Interlingua. The English word "wall" seems like a perfectly straightforward primitive concept to an English-only speaker. But when translating into French, we must distinguish between walls seen from the inside and walls seen from the outside. Will every new language we add to an Interlingua-based machine translation system force

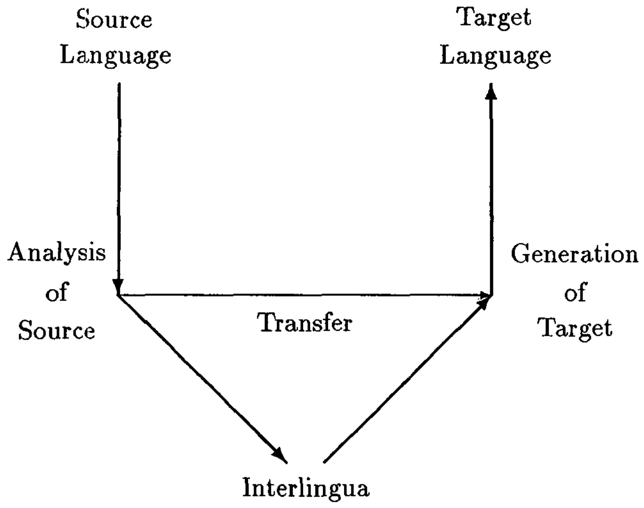

Figure 2: Interlingua Approach to Machine Translation.

on us further similar fragmentation of what we had believed were primitive concepts? The issue of how to construct an Interlingua incrementally is the topic of the paper by Hovy and Nirenburg.

Recently a new approach to machine translation—the Statistical Approach—has been attempted, principally at IBM Yorktown. In order for a statistical approach to work, there must be enough data available encoding the relevant information. This forces one to simplify the underlying model of the languages and the transfer between them to the point where statistical analysis becomes feasible.

In earlier work by Brown and his colleagues, the following simplifications were used. Instead of having a grammar of the source language, no analysis was done of the source text. Instead of having a grammar of the target language, a trigram model of the target language was used, capturing some but by no means all the structure of the target language. In place of transfer rules there were, first, a model of lexical correspondences between the two languages, and, second, a model of the alignment of corresponding sentences. The alignment model, rather than using the richer information about correspondences between parse trees that transfer rules seek to capture, encodes the simpler structural relations illustrated in Figure 3.

This approach is then illustrated in Figure 4.

There is no reason in principle that a Statistical Approach need be this shallow, and in fact in more recent work, the IBM Yorktown team has attempted to incorporate more linguistic structure into their efforts. As large bracketed corpora become available, it may be possible to align tree fragments as words are aligned today and derive transfer rules statistically.

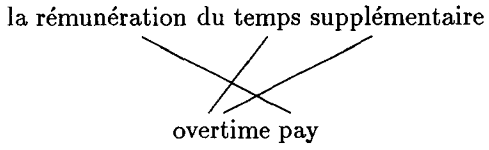

Figure 3: Alignment Information.

A problem that everyone who deals with real-world text eventually encounters is the problem of how to analyze very long sentences. This has especially engaged the attention of researchers in text understanding in the last several years. The solution is to break the sentences into phrases in the "right" way, whatever that is.

Statistical approaches to translation are computation-intensive, and hence sentences become very long a lot sooner than in other approaches. The problem of breaking sentences into phrases that can be translated independently must be faced a lot earlier. The paper by Brown et al. in this volume addresses this problem in the statistical translation framework.

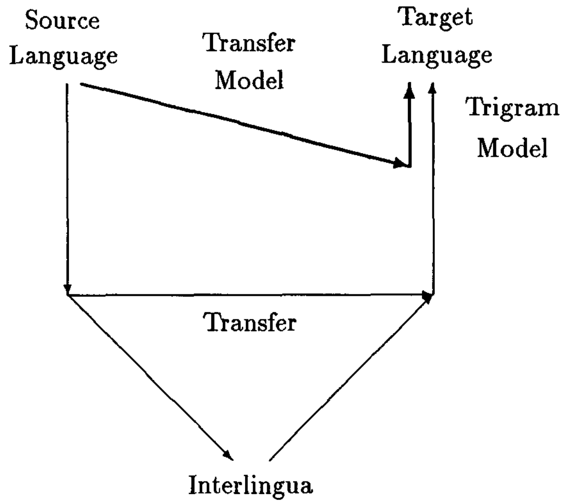

Figure 4: Statistical Approach to Machine Translation.