TOWARDS REVERSIBLE MT SYSTEMS

Pierre Isabelle

Canadian Workplace Automation Research Center
1575 Chomedey Blvd.
Laval, Quebec, Canada H7X 2X2

1. Why reversibility is an important issue

Translation competence is best viewed as a relation between an infinite collection of source language (SL) texts and an infinite collection of target language (TL) texts. Very few of the existing MT systems are based on an explicit characterization of the translation relation between pairs (or sets) of languages. Rather, most existing systems attempt to directly define a *particular one-way procedure* for mapping SL texts onto TL texts (or, perhaps, two separate one-way procedures L1 → L2 and L2 → L1).

In order to understand the difference, consider a simple example. In English-to-French translation, one can usually translate both the simple present tense and the present progressive tense as a French *présent simple*. From this perspective, the problem appears wonderfully simple, and the developers of English-to-French MT systems may well be tempted to forego any deeper analysis. However, the French-to-English perspective tells a different story. *Présent simple* does not translate indifferently as simple present or present progressive; it translates sometimes as one, sometimes as the other, according to various types of contextual indicators.

The case of the correspondence between French *tu* or *vous* and English *you* provides an example of the reverse situation: French-to-English translation appears straightforward while English-to-French translation appears problematic. The fact is that a directional perspective almost never tells the whole story.

One possible and common reaction is to conclude that directional systems are easier to build and constitute a more realistic goal for MT research. I do not subscribe to this interpretation. One's viewpoint determines how things look, not how things are. The conditions under which one chooses to translate *présent simple* as either simple present or present progressive directly pertain to the correct description of tenses *in English*. The present progressive is used in English with activity verbs in nongeneric sentences, when the activity extends across discourse time (the details need not concern us here).

These conditions are an integral part of English grammar. The fact that they are partly foreign to

French grammar obviously should not mean that one can safely exclude them from a grammar of English. Every time a grammar is weakened, increased *overgeneration* results. In practical terms, this means that the analysis component of the MT system will produce more and more *spurious ambiguities*. For example, a grammar of English that lacks whatever knowledge is required to filter out *He is knowing English* as a possible translation for *Il connaît l'anglais* is also unable to select the correct interpretation for *They are knowing readers.* Ultimately, this can only mean an increased number of mistranslations.

2. Reversibility: the practice

An MT system can only be said to be reversible if every one of its components is reversible. In particular, analysis and generation of a given language must be based on the same grammar and the same dictionary. In this respect, several proposals have recently been made in the litterature (e.g. Dymetman & Isabelle 1988, Shieber et al. 1989). And there exists at least one example of a reasonably sized fully reversible MT system (Isabelle et al, 1988).

From the developer's point of view, reversibility brings with it some important practical benefits. It is much easier to know how accurate a grammar really is when it can be used for both analysis and generation. Parsers are most often based on grammars that grossly overgenerate, while generators tend to be based on grammars that undergenerate. In a directional system, these shortcomings are hard to detect. Bidirectionality quickly makes obvious where the grammar goes wrong. The performance of a reversible translation system can be tested by the production of *retranslations*. If, for example, your system flatly equates both the simple present and the present progressive with the *présent simple*, the retranslations will promptly remind you that you tried for the easy way out but did not get where you wanted to.

Of course, the other side of the coin is that the reversibility test is a very tough one to meet. In fact, I don't believe that we have yet reached the point where we can get rid of the pervasive procedural bias that cripples large-scale operational MT systems. However, I am convinced that in order to make substantial progress, MT research must attempt to produce much more precise characterizations of the translation relation between pairs (or sets) of languages.

Dymetman M., Isabelle P. (1988) *Reversible Logic Grammars for Machine Translation*, Proceedings of the 2nd International Workshop in Machine Translation of Natural Languages, CMU, Pittburgh.

Isabelle P., Dymetman M., Macklovitch E. (1988) *CRITTER: a Translation System for Agricultural Market Reports*, Proceedings of COLING-88, Budapest.

Shieber S., van Noord G., Moore R., Pereira F. (1989) *A Semantic-Head-Driven Generation Algorithm for Unification-Based Formalisms*, Proceedings of ACL 89, Vancouver.