

FORUM ON MACHINE TRANSLATION

What Should Machine Translation Be?

John S. White
Siemens Information Systems
Linguistics Research Center
PO Box 7247 University Station
Austin, TX 78712

MODERATOR STATEMENT

After a considerable hiatus of interest and funding, machine translation has come in recent years to occupy a significant place in the discipline of natural language processing. It has also become one of the most visible representations of natural language processing to the outside world. Machine translation systems are relatively unique with respect to the extent of the coverage they attempt, and, correspondingly, the size of the grammatical and lexical corpora involved. Adding to this the complexity introduced by multiple language directions into the same system design (and the enormous procedural problems imposed by simultaneous development in several sites) gives some clue as to the optimism which presently exists for machine translation.

It is obviously believed in many quarters that computer science and linguistic science have become sufficient for production-environment machine translation. Private sector companies continue to introduce new MT systems to the marketplace worldwide, and many more are venturing into development and implementation. The industrial interest, meanwhile, has been instrumental in opening up possibilities for doing basic research in it, in part because of direct interaction between industry and research, and in part because of the overall increased awareness. It is indeed worth speculating whether renewed interest shown by governmental scientific agencies is related to the level of commercial acceptance.

But some feel that this visibility causes more harm than good. The concern has been expressed that an operational failure in machine translation will be seen as a failure in natural language processing generally, that a particular implementation rejected by users could cause a snowball ultimately resulting in the demise not just of MT as in the ALPAC aftermath, but also of all of computational linguistics.

Some may go so far as to suggest that such a day of reckoning will be inevitable as long as production-level machine translation efforts continue.

If it is indeed the case that production machine translation is not feasible, then machine translation is at best a heuristic environment for experimentation in linguistic theory. And machine translation does serve such an end admirably well: the modularity of program and linguistic description of which a well-designed translation system is capable allows work on hypotheses within one linguistic theory, or evaluation of different linguistic theories, without fundamental changes to the computing environment.

Two positions are identified here, whose distance from each other serves perhaps to encompass the whole range of thought on the ultimate potential of machine translation, as well as on the best possible design of a translating device. The one position holds that MT is a viable production tool whose benefit is more than worth the immense effort involved in linguistic description, textual coverage, and coordination of multi-national development. The other position holds that MT is a useful laboratory for linguistic study in a small, easily maintainable computing environment.

Despite the polarity, there is a common ground, which we employ as the datum point from which to explore the issues in machine translation today. We have progressed from the debate about the possibility of machine translation to the debate about what machine translation should be. This in itself is indicative of our awareness of the progress of computational linguistics as a whole.