

The Inflexible Fickleness of Fashion

Sergei Nirenburg

*Computing Research Lab
New Mexico State University
Las Cruces, NM 88003*

Machine translation has been a fashionable field for at least forty years of its fifty-year history. The reasons for this vary from R&D glory to commercial payoff. Over the years, an impressive variety of methods have been used as the basis for translation programs. The problem has, however, proved so complex that the quality of the final result has not correlated significantly with the method chosen. Rather, it correlated with the amount of descriptive work on language that was carried out.

Of course, MT research has brought about significant side benefits. Entire scientific fields were created largely due to MT efforts: witness the nascence of computational linguistics. Often, MT was used as an application of choice for a variety of workers to test and attempt to corroborate their theories of language and of human thinking capacity. It is characteristic that the final report of the Eurotra project listed as its major success the creation of computational-linguistic infrastructure in the countries of the European Community deemphasizing the fact that no realistic MT system was built under its auspices. Many factors contributed to the lack of the engineering achievement in this project, among them the relative lack of accent in Eurotra on actual description and system building, with preference given to designing detailed formal specifications of (largely syntactic) levels of analysis and their corresponding formalisms.

Is the Eurotra case prototypical for the entire field of MT? One of the problems with the field has been that the descriptive work is, frankly, rather monotonous and boring. This is why attempts were made either to make it less boring (by adding an independently motivated theoretical angle to the descriptive work) or to try to avoid it altogether.

The latter objective was made manifest in a) attempts to use AI learning techniques or more practical semi-automatic procedures for knowledge acquisition and b) the application of statistical methods for establishing cross-linguistic correspondences in lieu of language description work. The former solution made itself manifest in viewing MT as a testbed for one's favorite linguistic or computational-linguistic theories, such as, for instance, the currently fashionable "principle-based" approach to syntax. Machine translation is indeed a tempting avenue of computational inquiry into modeling human mental and language processes, and a number of approaches to NLP in AI dabbled in MT as a potential application. Knowledge-based MT is a direct offshoot of the AI tradition.

The most remarkable feature of the statistical methods in MT is that they are not at all specific to their subject matter --- the same techniques applied to processing language could and are used, for example, in the studies of the human genome.

The current R&D-oriented MT approaches, whether rule-based or statistical or hybrid, are based on “imported” ideas. At the same time, the best systems on the market cannot boast much by way of technological or scientific advances. Instead, they rely on brawn: huge, handcrafted dictionaries and grammars and a plethora of specialized translation routines. All of us are curious to see how well the R&D approaches will work once sufficient resources are allocated for one or more of them to reach the status of a product. The question is: what kind of imported techniques shows the most promise? The answer is not clearly obvious and is determined by sociological (read: the vagaries of funding) as well as scientific and technological trends.

The major scientific (or methodological) trend in the field is experimenting with how well the statistics-oriented methods will advance the state of the art in MT without the need for massive manual knowledge acquisition.

The major technological trend in the field looking for the best ways of mixing the statistical and the “rule-based” methods. This author has been an early advocate of mixing such methods at the level of their final results, a method called multi-engine MT. Other approaches seek a more involved interaction, with statistics used not only during the process of MT but also to support development of background resources (i.e., dictionaries and grammars).

The major sociological trend, at least in the US, is the emphasis on a regimen of evaluations and competitions among MT (and, more broadly, NLP) systems. This promotes rigor and discipline as well as conformity and search for local solutions which are not necessarily the most promising ones in the long run. Approaches that show a steady improvement are rewarded. Approaches with long gestation periods are punished.

Emphasis on mixed approaches is, for non-statisticians, a rearguard regrouping action, while for statisticians (witness the evolution of the claims and practices of the Candide IBM MT group), a search for any avenue for improving the rather modest final results.

The knowledge-based and linguistics-based methods will do good by regrouping and concentrating on those tasks and situations in which statistical approaches fail to deliver. One must, however, remember the lesson of computer chess: at present, the best chess-playing systems are not terribly knowledgeable about chess strategy and tactics but they consistently beat AI-based programs and compete on equal terms with grandmasters. The \$64,000 question is: how much more complex is human translation ability compared to the human chess-playing ability? That is, for how long will there be an opportunity to study language use through MT? If statistical methods succeed, rule-based MT may go the way of the AI-based chess programs.

My personal opinion is that MT is too complex to be fully accounted for by the current statistical processing methods, even though these methods do not aspire to building representational models of human language capacity and rely only on the input-output behavior of such models (in MT, a text and its translation). In the final analysis, the open-endedness of language will become the stumbling block for these methods conceptually, just as, logistically, the chronic shortage of resources (bilingual corpora) may precipitate the swing of the pendulum of MT R&D fashion back to the mentalist camp from its current behaviorist direction.

How long will this take? If history is any guide, such swings come roughly every 30 years: mentalism was in scientific ascendancy between 1960 and 1990, while behaviorism reigned, at least in the US, for about thirty years prior to that. Of course, we cannot be certain that we are witnessing this pendulum swing and not some other, unconnected development. Time will show. A more intriguing thought is that, just possibly, the rule-based/corpus-based dichotomy is not as important as we currently think. Maybe the real problem of MT as technology is that it is not generally understood how difficult the problem actually is. The confident claims, made by newcomers to MT (including this author some fifteen years ago), help stoke the high expectations of getting the desired result with a modest expenditure. At the current level of MT R&D, either the expectations should be lowered or the time scale of getting the results must be significantly extended. For best results now, it might be necessary to fund a language description effort of truly Tower-of-Babel proportions.