

# MT and Language: Conflicting Technologies?

## Ariadne's Endless Thread

Alex Gross

In a previous piece (*Where Do Translators Fit Into Machine Translation?*), I sought to direct a variety of philosophical, linguistic, and practical questions to members of the MT community during one of their major international conferences. Since response to these questions has been less than deafening, I would now like to suggest a few possible answers and speculations of my own concerning these matters. Some bitterness has crept into MT discussions of late, and so I would like to emphasize once again that no reasonable person is opposed to MT where it works. The question is a more theoretical one, though rich in practical applications, and concerns how far MT is truly capable of improvement and why it has taken so long to reach its present condition. In this discussion I propose to deal with both MT and human language as specific "technologies," an approach as obvious for the former as it may seem surprising for the latter.

It is not at all hard to show that MT comprises some sort of technology. The reduction of knowledge to bits and bytes, the building of algorithms, the construction of programs are all processes familiar to us from other branches of computer technology. And indeed MT was foreseen from the beginning by such computer pioneers as Turing, Shannon, and Weaver as a rich potential application. Even in commercial and practical terms, MT would appear at first glance to have passed through all the usual stages common to technologies:

1. Need (or perceived need).
2. Determination of technological feasibility.
3. Successful financing.
4. Basic research and development.
5. Preparation and testing of prototypes.
6. Further improvements and developments.
7. Launching of commercial products.
8. Publicity and marketing.
9. Operator or consumer training in their use.

Nonetheless, a closer examination of these stages reveals several points at which MT may have already fallen short. It can be argued, for instance, that the "need or perceived need" for MT was never sufficiently demonstrated, as no trustworthy figures have ever existed concerning the actual or potential total world volume of materials needing translation nor of the number or capabilities of human translators ready to translate them, nor—finally—of the real or potential economic benefits to be reaped from introducing this new method.

Further reservations may be expressed concerning the basic "research and development" process out of which MT has grown. Essentially all "computational linguistics" has been based in or grown out of the prior theorizing of conventional linguistics. But for some decades the study of linguistics, never a rigorous science to begin with (despite some efforts to make it one), has been subject to a process of growing decadence and obfuscation. This process has gone so far that departments of Linguistics have recently been disbanded at two major universities, and many scholars now regard the field as even less respectable than sociology.

Further discussion of the linguistic side will be postponed until we have had a chance to consider whether and, if so, how language itself may be considered to be a technology.

Further objections as to how well MT has lived up to three other stages in our profile—namely, launching of commercial products, publicity and marketing, and operator or consumer training—can also be voiced, but this matter will also be overlooked for the time being.

There are of course other computer-specific steps in developing a technology—such as reverse engineering pre-existing programs or the use of orphan code—which have helped to speed up the development of applications in the past, and in most fields we have also witnessed the effects of economies of scale. It is partly due to these last that we have seen calculators shrink from desktop giants to the size of visiting cards within our own lifetimes. Comparable developments in other fields have led many to suppose that virtually anything is possible.

At this point it is also important to note that MT is most definitely—and perhaps most self-definingly—a component part of AI, or Artificial Intelligence. Certainly the AI Community has done all within its power to encourage funding sources and the general public to believe that computers can do almost anything. While MT advocates now concede—at least among translators—that FAHQT (Fully Automatic High Quality Translation) may never happen, the AI Community at large has never made any such concession. On the contrary, at a recent conference its so-called HAL wing proclaimed its allegiance to recreating full human intelligence—including language comprehension—within a computer. This is not surprising news to those who have lurked on Internet's comp.ai newsgroup. FAHQT would of course be a relatively simple task for such a computer, assuming it could be built.

Now that we have seen how MT conforms—with some apparent exceptions—to the overall pattern of a technology, let us next examine the qualifications of human language in this regard. It is obvious from the beginning that any such claims will have to be expressed in biological and physiological terms, since human language did not develop in the same way as technologies such as metallurgy or computer science, even though the latter are arguably its offshoots.

The long-debated origins of language—variously attributed to the "Bow-Wow Theory," the "Yo-Heave-Ho theory," or the "Pooh-Pooh Theory"—are so inauspicious and unpersuasive that readers may wonder what point there can be—like so much else in linguistics—to any further discussion at all. But once we turn our attention to biological development, both of the species and of our related animal cousins, a different perspective may unfold, and some startling insights may just be within our view. As human beings we frequently congratulate ourselves as the only species to have evolved true language, leaving to one side the rudimentary sounds of other creatures or the dance motions of bees. It may just be that we have been missing something.

On countless occasions TV nature programs have treated us to the sight of various sleek, furry, or spiny creatures busily spraying the foliage or tree trunks around them with their own personal scent. And we have also heard omniscient narrators inform us that the purpose of this spray is to mark the creature's territory against competitors, fend off predators, and/or attract mates. And we have also seen the face-offs, battles, retreats, and matings that these spray marks have incited.

In an evolutionary perspective covering all species and ranging through millions of years, it has been abundantly shown time and time again—as tails recede, stomachs develop second and third chambers, and reproduction methods proliferate—that a function working in one way for one species may come to work quite differently in another. Is it really too absurd to suggest that over a period of a few million years the spraying mechanism

common to so many mammals, employing relatively small posterior muscles and little brain power, may have wandered off and found its place within a single species, which chose to use larger muscles located in the head and lungs, guiding them with a vast portion of its brain?

This is not to demean human speech to the level of mere animal sprayings or to suggest that language does not also possess other more abstract properties. But would not such an evolution explain much about how human beings still use language today? Do we really require "scientific" evidence for such an assertion, when so many proofs lie self-evidently all around us? One proof is that human beings do not normally use their nether glands to spray a fine scent on their surroundings, assuming they could do so through their clothing. They do, however, undeniably talk at and about everything, real or imagined. It is also clear that speech bears a remarkable resemblance to spray, so much so that it is sometimes necessary to stand at a distance from some interlocutors.(1)

Would not such an evolution aptly explain the attitudes of many "literal-minded" people, who insist on a single interpretation of specific words, even when it is patiently explained to them that their interpretation is case-dependent or simply invalid? Does it not clarify why many misunderstandings fester into outright conflicts, even physical confrontations? Assuming the roots of language lie in territoriality, would this not also go some distance towards clarifying some of the causes of border disputes, even of wars? Perhaps most important of all, does such a development not provide a physiological basis for some of the differences between languages, which themselves have become secondary causes in separating peoples? Would it not also permit us to see different languages as exclusive and proprietary techniques of spraying, according to different "nozzle apertures," "colors," or viscosity of spray? Could it conceivably shed some light on the fanaticism of various forms of religious, political, or social fundamentalisms? Might it even explain the bitterness of some scholarly feuding?

Of course there is more to language than spray, as the species has sought to demonstrate, at least in more recent times, by attempting to preserve a record of their sprayings in other media, such as stone carvings, clay imprints, knottings in beads, and of course scratchings on tree barks, papyri, and different grades of paper, using a variety of notations based on characters, syllabaries or alphabets, the totality of this quest being known as "writing." These strivings have in turn led to the development of a variety of knowledge systems, almost bewildering in their number through various eras and cultures in a multi-dimensional, quasi-fractal continuum. Thus, language may turn out to be something we have created not as a mere generation or nation, not even as a species, but in Von Baer's sense as an entire evolutionary phylogeny. It is this greater configuration which may transcend the more primitive side of language and eventually provide a more complete image of its nature, perhaps even shedding light as well on the nature of human knowledge itself.

In the face of this imposing prospect, it is not surprising that MT advocates almost invariably focus on that part of language devoted to "verbal meaning." But I have listed elsewhere no less than five other common functions of language, almost none of them totally devoted to the communication of verbal meaning. They are as follows:

1. Demonstrating one's class status to the person one is speaking or writing to.
2. Simply venting one's emotions, with no real communication intended.
3. Establishing non-hostile intent with strangers, or simply passing time with them.
4. Telling jokes.
5. Engaging in non-communication by intentional or accidental ambiguity, sometimes also called 'telling lies.'

6, 7, 8, *etc.* Two or more of the above (including communication) at once. (2)

It should be obvious that most of the foregoing conform at least as well to the model of "spraying one's surroundings" as they do to communicating verbal meaning as such. It is hard to see how MT can ever hope to cope with these larger problems, and it is not surprising that we have recently seen various limitations arise connected with launching, marketing and publicizing commercial MT products as well as with training translators to deal with MT output as post- editors.(3)

Under no circumstances is this "spraying" metaphor being presented as a total account of language. This aspect is considered quite briefly—among many other intellectually more respectable analogies for language—in the forthcoming ATA Scholarly Volume on Terminology, and the author hopes to provide an even more rounded account in a work still being completed. It does seem important, however, that some relatively primitivist footnote to the origins of language should be introduced into discussions about linguistics and its applications, MT among them. Much writing about language—since it is scarcely uneducated people who write about this subject to begin with—tends to luxuriate in self-importance and self-congratulation about how important a development language has been for humanity. But the rational and intellectual aspects of language are in a sense only the most obvious ones, which may have led MT advocates, perhaps following Chomsky, to suppose language possesses a logical substructure it may in many cases actually lack.

Contrasted with these more complex aspects of language, a good computer program should be a model of simplicity. It should solve its problem in the most elegant way and—as though following the thread of Ariadne—it should go directly to its goal and craftily find its way out of the labyrinth again, easily slaying or avoiding all minotaurs and monsters along the way and using its thread as a guide rather than tripping over it as an obstacle. If it must double back occasionally in its path, there are good and cogent rules for not letting this prove a distraction. It is thus not surprising that the labyrinth or maze is an image that finds instinctive resonance among hackers, nor that they take delight in playing games where monsters must be slain.

But what computer rules will guide us through the labyrinth of language? There is no one entrance or exit and no definable center. We have all had to learn this labyrinth step by step simply to come as far as we have. We have even learned about the computer—up to a fairly advanced point—mainly by using language. When we try to solve the problems of language, whether by building MT programs or Voice-Writers or other Natural Language applications, we suddenly find there are monsters everywhere, and it is they who slay us, rather than the reverse. The technique for slaying one language monster may allow another to triumph. And the thread itself no longer traces a brief or elegant path, it has in fact become endless in its back-trackings and recrossings, creating a whole new jungle of Koenigsberg Bridges, Towers of Hanoi, Traveling Salesman's Problems, and other computer math anomalies. Worst of all, the labyrinth of language is not some separate location we can visit at our convenience and slowly come to know. Rather, we have no choice but to live in it constantly. We have never lived anywhere else.

Perhaps it is time to glance backwards from a systems perspective and see how well language has conformed to our nine-point profile for a technology. Clearly no survey of need or technological feasibility can have taken place in the conventional sense. Nor was financing or research and development a major factor, since a whole succession of species was available as a free laboratory over several million years. But at the right time, language came to be installed in the entire human race, at first only spoken but finally written as

well. It was clearly a technological advance, since it made it possible for humans, even in its oral form, to exchange more complex observations and measurements than could be passed along without it. Perhaps most impressive of all, language now has a total installed base of over five billion living systems, something no computer can remotely match, and is still expanding. Its one main drawback as a technology may lie in the huge service and administrative staff of teachers, writers, editors, and critics needed to maintain it, though a comparable problem is not unknown with computers.

At computer conferences one frequently hears programmers and other specialists complaining about natural language and boasting about how they live in a purer, more perfect sphere, in a truer reality, whether virtual or otherwise. One day they will supplant all the confusing skeins of messy reality and even messier language with a finer, higher, texture of purest logic, and all the world will instantly evolve to the next more transcendent stage. Those who voice these boasts have but a single problem: for the time being at least, they are forced to express their vision in precisely the natural language they claim to despise. To perfect MT or any natural language application, there is no escaping the fact that it will be necessary to build a language both higher and lower, in human and computer terms respectively, than the one we now use, a true metalanguage. There is room for a great deal of skepticism as to whether this is possible.

I am not so sanguine as to hope that the foregoing will have any effect at all on MT zealots, Hal AI acolytes, or dedicated programmers.(4) Like heroes of old intent on slaying the foe at any cost, they pay heed only to news of the latest new weapon alleged to have power against the minotaur. It may be called Corpus-Based MT, or Neural Nets, or Hidden Markov Models, or Three-Dimensional Fuzzy Logic, or perhaps it may hinge on creating a neurological interface with the brain itself. Or it may simply be a matter of time— after all, when computers become sufficiently large and inexpensive, nothing will be beyond their power, or so goes the tale. But without a complete algorithm for handling language and linguistic problems, not all the power in the universe can withstand the might of the great God GIGO: Garbage In, Garbage Out.

Some of these approaches may bring some advances to some aspects of MT. But programmers, AI enthusiasts, and MT researchers alike would do well to realize that they too live in the labyrinth of language, a realm whose navigational problems have long been underestimated.

#### NOTES:

(1) This resemblance extends even to the etymology of the two words, speech and spray, which are closely related in the Indo- European family, as are a variety of words beginning with "spr-" or "sp-" related to spraying and spreading: English/German *spread, sprawl, spray, sprinkle, sp(r)eaK, spit, spurt, spout, Spreu, spritzen, Sprudel, Spucke, spruehen, sprechen*, Dutch *spreken*, Italian *sprazzo, spruzzo*, Latin, *spargo*, Ancient Greek *spendo, speiro*, etc. The presence of the mouth radical in the Chinese characters for "spurt," "spit," "language," and "speak" also shows how related these concepts are on a cross-cultural level.

(2) From the author's *The Limitations of Computers As Translation Tools*, a chapter from *Computers in Translation: A Practical Appraisal*, edited by John Newton, Routledge, London, 1992.

(3) Peter Wheeler: *On Using Professional Translators to Post-Edit*, pp. 353-59, *Looking*

*Ahead, Proceedings of the 31st Annual Conference of the American Translators Association*, Edited by A. Leslie Willson, Learned Information, Inc, 1990.

(4) I wish there were some way both programmers and translators could become aware of their many similarities. Both work at extremely demanding intellectual tasks requiring a high level of familiarity with specialized knowledge. Both tend to live somewhat solitary lives, punctuated by moments of self-indulgence. Both are beset by constant deadlines, and both are reputed to be something of drones. While the programmer often purports to despise language and sees himself as living in "Cyberspace," the translator may feel hostile towards computer logic while setting up an almost mystical relationship with his dictionaries and envisioning himself as dwelling in a realm where reality and meaning meet. Perhaps both are mistaken in somewhat similar ways.