Working toward Success in Machine Translation

Laurie Gerber

Language Technology Broker
4774 Del Mar Avenue
San Diego, California, USA

LTB@gerbersite.com

Abstract

In this short paper, I explore ways in which the MT community might formulate goals that will expand on known successes, build on existing strengths, and identify long term research goals.

Introduction

While developments in MT technology and theory have continued to chip away at the problems of MT, it is not always clear what we are chipping away at. There are problems raised within various theoretical frameworks. There are the general problems of raising the quality ceiling, or reducing the cost of reaching the inevitable quality ceiling. But do the solutions or performance improvements represent milestones of progress? On the path to where?

In the 1920s, mathematician David Hilbert posed 23 problems to the mathematical research community, some of which have been solved, and some of which continue to occupy the mathematical community. A problem for MT researchers and funders is the absence of a core set of issues to be addressed. The problem is exacerbated by the diversity of approaches to MT currently available. In our desire to be neutral with respect to approach, we may be even less able to come up with a set of core issues that everyone can agree on. However, given the rather low quality ceiling we encounter for general-purpose MT, regardless of approach, there is powerful motivation to come to a consensus on where we need to invest our effort. We can only hope that through forums like MT 2010, a set of problems can be identified which will fortify the MT community by defining a set of problems, and ignite the funding communities by giving clear definition to what should be accomplished.

I suggest two areas in which such problems can be clearly formulated and their solutions clearly assessed. The first relates to applications, and is purely practical and relatively near term, aimed at helping MT to "earn its dinner" while it grows up. The second suggests a number of basic goals for the research community.

Practical Applications

Machine translation is not an easy technology to love. It is not much of a growth industry, and while new developers pop up with what they hope will be revolutionary ideas, most of these enterprises eventually give up or settle into the subsistence business of producing working MT systems. The reason ALPAC didn't kill MT entirely, and that these subsistence businesses survive, is that there are still a number of applications for which MT, even in its current state, is the best or only solution.

Let's call these applications the success stories for MT. What is striking about these success stories is that they all contain two common elements: 1) Some person in the user or service provider organization who has the commitment and resilience to persevere through the difficulties and expense of implementation. 2) An application that can benefit from the strengths of MT, and tolerate its weaknesses.

The first one is clear to anyone familiar with MT implementations. Concerning the second one, such applications can be characterized by how they affect the process into which they are integrated. The point is that translation is only a worthwhile activity if having the translation makes something else possible. For example, why do intelligence gathering organizations do translation? Because it gives them access to information. MT crucially makes that access more timely. Why does a company like Caterpillar do translation? To extend their business internationally. MT allows nearly simultaneous release of products internationally. The point I am making is a slightly different one from the more general categories of assimilation and dissemination, under which these uses of MT may be subsumed. One can imagine many motivations for assimilation and dissemination of information. My slant concerns how we understand when and where MT can accelerate or enable key business processes. Are MT developers in the translation business? No. When they are providing MT to intelligence-gathering groups, they provide efficient **access**. When they are providing MT to businesses with international sales, they provide an **efficiency** tool that accelerates the business cycle, and gives a competitive advantage. New applications are emerging for which MT is a similarly enabling technology: real-time website translation – for example by news services, or by companies seeking to reach new markets with products they advertise online. Neither fish nor fowl, this application looks like assimilation to its users (consumers of the translated information), especially in the case of news services. But it looks like dissemination from the information provider's side. It provides **access to new markets for information.** In the case I am considering, it deals with time-sensitive information that must be retranslated after every update to the source, into multiple languages. It is an

application that was not feasible with manual translation and hence, it does not replace translators, it replaces non-translation. In 1993, (Church and Hovy) suggested some "Good applications for crummy machine translation". We need a systematic follow up that characterizes the success stories in more rigorous terms than I have done here. It may be that until high quality MT comes along, regular use of MT will be limited to the odd organization that has the vision and determination to overcome the obstacles. However, it may also be possible to move from a subsistence business to growth industry with better focus on, and development that targets, the applications where MT provides a unique solution. Pitting MT in its current state against human translators is a poor idea (Kay '73). The best recent evidence is the way that translation memory has edged out MT in the serious translation communities as a solution to the need to increase productivity. We need a better understanding of how and where these ideal applications for MT are.

Generating more success stories

Earlier, I mentioned that the success stories were all applications that were able to make use of the strengths of existing MT systems, and were tolerant of their weaknesses. One typical path to success is specialization. For various reasons, though, funding and development of MT has tended to focus on general purpose systems. Since we don't yet know how to produce high quality general purpose systems, why not see how far we can go with specialized systems? For example just for email, or just for business letters, or just for patents (as with Pa-Trans). Usability is much higher, and development costs are lower. Even if we can't tackle general purpose MT head-on, Maybe we can approach general purpose inductively, as we understand more about what is involved in various specialized translation tasks.

Theoretical Directions

Measuring text to measure transfer of meaning

In spite of all the effort spent on developing and refining extrinsic and intrinsic evaluation methods for MT, we will not really be able to measure the quality of MT until we can measure something much more basic – the content of a text. How can we measure a conversion process if we cannot measure, or even adequately characterize, the before or after states? The ability to fully characterize texts, so naturally done by any human reader, is fundamental to any meaning-based text conversion process, including translation and summarization. Full characterization should include the real-world state of affairs implied by each part of the text, and the communicative goal/effect/tone of the text. This information from a text is more important than the compositional structure of the text, and certainly more important than the discourse or syntactic structures of its components. Surely the inability to fully characterize a text is behind poor human translation as much as it is behind poor machine translation. Until we can predictably and repeatably perform such analysis and representation of monolingual texts automatically (which presupposes the development of a repeatable,

verifiable manual process first), we really can't hope for high quality automatic translation. I don't hear people refer to the acronym FAHQMT much, except as an old and quixotic notion that is part of the lore of MT past. And yet, I believe that the hope of producing high-quality translations, even communicating machines, is still a driving force in our efforts at creating human language technologies. The goal of a complete representation of text is clearly not part of any plan that will be realized in 2010. But the absence of such long term goals makes it difficult to plot a meaningful path to the future.

Understanding limitations

Thinkers on language have described why translation is difficult, and why something like fully automatic general purpose MT may not be a reasonable goal (Melby 1995). And yet there are still compelling reasons to continue to pursue fully automatic translation. Various approaches have had success in advancing the quality of output, or improving the efficiency with which MT systems are developed, but all encounter a performance ceiling. We need a better understanding of exactly what constitutes those performance ceilings, and how the performance limitations differ between approaches. Many MT development methods are expensive and time consuming. The investment just to get to the point of encountering performance ceilings is tremendous. In order to optimize the effort, particularly in research systems where public funds (and public opinion) are at stake, we need a way to understand and employ best practices. There may be many factors behind performance limitations, including the lack of world knowledge and poor coreference processing, but at this point we should be able to be a bit more precise about what makes up the 10% or 20% or 50% that wasn't translated correctly in a 90% or 80% or 50% accurate translation.

Understanding tolerances

Human readers make good use of badly written and translated text all the time. The difference between such human-produced defective texts and machine translated defective texts is generally the type of error that is made. The ideals of machine performance hold machines to the very highest levels of human performance. Since it is unlikely that we will be able to accomplish every goal with respect to improving MT within the next 10 years, efforts should focus on the issues that most affect usability. This may require research programs to defer work on the relatively superficial issues that readers can tolerate. It will also require investigation into the nature and degree of those tolerances, or as a simpler approach, task-based MT evaluation keeps attention on the big issues, rather than focusing on superficial lapses that may seem to dominate purely quality oriented evaluations.

Build on strength

Two points from the known success stories suggest directions for research that would build on strengths:
First, systems for controlled languages, or for sublanguages, have given us some of the nicest success

stories in the history of MT, where users actually get high quality output to work with. Rather than being a disappointing limitation of MT, these success stories suggest a direction for MT research. Such specialized applications succeed because they try to model a subset of the language, including its lexicon and structure. General purpose applications fall down partly because they try to ignore differences of genre, register, and subject domain. Human translators are considered expert only if they can accurately transfer the pragmatics, as well as the information content, of a text from one language to another. The best human translators are highly specialized with respect to subject domain and flexible with respect to genre and register. The research direction I propose here is one that seeks to develop dynamic, efficient, multi-dimensional models of language which can be reflected in both a lexicon and grammar (and possibly discourse structure). Some work along these lines has been done for a number of grammatical properties of various genres and registers of English text (Biber 1986). Such work needs to be extended to discourse and the lexicon. And by "lexicon" I mean not just the words and terms that are used, but the ways in which argument structures vary as well.

Second, MT has generally done best with technical text, in which the social and contextual aspects of communication are least present. That is, such texts are designed to be primarily informative. Translation of such language can be treated much like an encoding, unlike general text which introduces social and contextual factors. It would be helpful to develop suites of text that represent a continuum from the least to most situated (context-sensitive) types of text. Successfully translating highly formal text is a step that would presumably need to preceed translation of increasingly context-sensitive texts. By approaching translation of this aspect of textual complexity as a series of increasingly difficult challenges, we may be better able to make sense of the real capabilities and sophistication of MT engines.

Don't build toy systems

Bernard Scott, on a panel discussing MT research at AMTA 1994, suggested that toy systems demonstrate nothing about the nature of MT, because virtually the whole problem for MT is scaling up. This is not to say that specialized systems say nothing, but that systems which are only prepared to handle a small set of filtered inputs tell us nothing interesting about how to do MT in the real world.

Move beyond the sentence

Experiments by Daniel Marcu (Marcu et. al. 2000) have shown how much coherence and understandability are lost when the sentences of an English text are rearranged, as well as showing the difference in discourse structure between English and Japanese text. Harold Somers has often described human translation strategy as "structure preserving as the last resort" whereas MT is structure preserving as the primary strategy. The MT system of the future has to take into account the whole structure of text, rather than

perennially trying to see how much it can get away with in ignoring discourse and text structure.

Conclusion

While the history of machine translation is not one of huge general success, there are many small successes to build on. As demand mounts for high speed translation, the MT community desperately needs a realistic roadmap toward the future. In this paper, I have suggested a few practical steps to expanding the profitability and perceived success of MT in its current state, as well as some strategies that work towards new levels of quality and sophistication, building on known strengths wherever possible.

Bibliographical References

Biber, D. (1988). Variations Across Speech and Writing. Cambridge University Press.

Church, K.W. and E.H. Hovy. (1993). Good Applications for Crummy Machine Translation. Machine Translation 8 (239--258).

Kay, M. (1980). The proper place of men and machines in language translation. Machine Translation, Kluwer, Netherlands 12, 323. Xerox PARC Working Paper, 1980.

Marcu, D., Lynn Carlson, and Maki Watanabe (2000). The Automatic Translation of Discourse Structures. The 1st Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL'2000), Seattle, Washington, April 30-May 4, 2000.

McCormack, Mark. (1988). What They Don't Teach You at Harvard Business School. Bantam Doubleday.

Melby, Alan, with C. Terry Warner. (1995). The Possibility of Language. John Benjamins. Amsterdam/Philadelphia.