<div align="center">**Panel**</div>

<div align="center"># The Limits of Automation: Optimists vs Skeptics</div>

For building resources and performing MT: Do automated techniques deliver? This (possibly controversial) panel will examine where the recently developed techniques of automated (statistical and other) methods for MT and Computational Linguistics are leading. In addition to performing some subtasks of MT, these techniques have proven rather useful for building resources. But, ultimately, the question remains: how much do these new techniques help? Do system builders still have to go in there by hand, to enrich the lexicons and fix the rules? Is it a matter of now being able to come up to speed more quickly with a new system, or is there a true improvement in quality as well? In other words: do automated techniques deliver?

Panel members:
Eduard Hovy, USC/Information Sciences Institute (moderator)
Ken Church, AT&T
Denis Gachot, Systran
Marge Leon, Pan American Health Organization
Alan Melby, Brigham Young University
Sergei Nirenburg, NMSU
Yorick Wilks, University of Sheffield

Optimists and Skeptics

Kenneth Ward Church

AT&T Labs-Research

In addition to performing some subtasks of MT, these techniques have proven rather useful for building resources. But, ultimately, the question remains: how much do these new techniques help? Do system builders still have to go in there by hand, to enrich the lexicons and fix the rules? Is it a matter of now being able to come up to speed more quickly with a new system, or is there a true improvement in quality as well? In other words: do automated techniques deliver?

We all know what the title of this panel is really referring to. The optimists are the folks behind Candide and the skeptics are the folks behind Pangloss. So, we're back to the old debate between Candide and Pangloss.

The last time we had this debate, we used *rationalism* as the code-word for Pangloss and *empiricism* as the code-word for Candide. I generally try to argue that both positions are reasonable. It depends on what you are trying to do. If you are trying to explain all and only the agreement facts of human languages, as Chomsky was, then his position makes lots of sense. But on the other hand, if you are trying to reduce entropy over noisy channels, as Shannon was, then his position makes lots of sense.

	Rationalism	Empiricism
Advocates	Chomsky, Minsky	Shannon, Skinner, Firth, Harris
Model	Competence Model	Noisy Channel Model
Contexts	Phrase Structure	Ngrams
Goals	All & Only Explanatory Theoretical Perfection	Central & Typical Descriptive Applied Minimal Error (H)
Linguistic Generalizations	Agreement, Wh-movement	Collocations, Word Associations
Parsing	Principle Based	Preference Based
Strategies	CKY (Chart), ATNs, Unification	Forward-Backward, Inside-Outside
Applications:	Understanding Who did what to whom	Recognition Noisy Channel applications

In this case, it isn't clear whether machine translation is more like the tasks that Chomsky was interested in, or more like the tasks that Shannon was interested in. Ultimately, almost everyone ends up adopting some sort of hybrid position.

Wilks (1994), a skeptic if there ever was one, has criticized the optimists as "Stone Soup." His main objection is that they've been adding intuition to their statistics, and in the end, they will end up with mostly intuition and the statistics is just the stone in the stone soup. But I wouldn't go as far as Wilks has, in suggesting that the statistics don't do nothing. I was say that by adding intuition, Candide has been moving toward a hybrid. The hybrid position is unavoidable. Even Wilks (1994) ends up advocating a hybrid. You just can avoid it.

Of course, the details of the hybrid position vary considerably from author to author. I generally argue for a hybrid that looks a lot like Tukey's notion of Exploratory Data Analysis (EDA), a reaction to traditional hypothesis testing. Instead of using statistics to confirm or reject a pre-conceived hypothesis (or intuition), EDA advocates the use of statistics in helping to suggest hypotheses. EDA is much less radical than the self-organizing position, where it is suggested that statistics can do it all. EDA merely suggests that statistics can do more than just hypothesis testing.

So how does all of this debating affect the bottom line? If the only line we care about is FAHQMT, then it is all totally irrelevant, because no one is getting anywhere, as far as I can see. If we are willing to count spin-off resources like aligned corpora and spin-off applications like translation workstations with translation memories and just-in-time terminology features, then the optimists have done quite well. It is clear that the panel statement was carefully crafted to avoid discussion of spin-offs, but I'd rather ignore the panel statement, than end up concluding that no one is getting anywhere.

References

Wilks, Y. 1994. Stone Soup and the French Room. In Zampolli, Calzolari, Palmer (eds.) *Current Issues in Computational Linguistics: In Honour of Don Walker* (585-595).

Assimilation or Dissemination? That is the Question

Denis Gachot
SYSTRAN

The purpose of machine translation systems is to enhance human communication by breaking the language barrier. However, this apparently simple aim has proven most difficult and extremely challenging throughout our industry's history. Today, MT systems are being used by companies essentially for either the rapid assimilation of foreign information, or for dissemination of information on products and services to foreign customers. It should be noted that, so far, MT systems have been successful in their assimilation role, but have encountered strong resistance from professional translators when used in a dissemination mode.

In fact, most of MT's bad reputation in the public eye comes from its accuracy shortcomings (The spirit is willing...). Although translation accuracy is of importance, productivity, readability, and comprehension also need to be considered. We contend that MT systems already do bring increases in productivity, but there are also numerous ways to enhance them, and which techniques to use depends on what mode you intend to employ.

Assimilation

Users' accuracy expectations in this mode vary greatly from the foxhole setting (Is this document a cooking recipe, or about land mine location?), to the scientific researcher needing to understand the content and nature of various texts. MT systems dealing in this mode are fully automatic, and need to be able to process a large variety of texts covering numerous domains. New techniques, such as large scale corpus analysis, and context free parsers have increased MT system development speed. Also, the availability of free resources on the Internet (glossary, corpus...etc.) have greatly contributed. The time to develop a usable system for assimilation is only a fraction of what it was a few years ago. These recent developments now enable MT developers to secure funding on new language combinations (e.g., Serbo-Croatian into English) with the ability to produce useful systems in short order.

Dissemination

In this mode, MT systems are judged on their ability to increase productivity. The translation phase is only one step of the production of a foreign document. In spite of translator resistance, two characteristics of dissemination should facilitate the use of MT: the typically limited text domain, and the possibility to access all phases of production, from authoring to post-editing of translations. Developers can create tools for any of the following phases:

Before MT: Grammar checkers geared toward known MT shortcomings (e.g., flagging of homographs, ambiguous -ing forms...etc.), as well as the use of translation memory systems to increase output quality.

During MT: Text mining, using statistical tools to locate frequent word association (a must in technical English), in conjunction with user friendly dictionary building tools, allows rapid dictionary development. The user option to specify document type (Journalism, meeting minutes, parts list...etc.) can also boost parser performance.

After MT: The application of stand-alone target language generation with the aim of increasing readability, as well as the use of specialized post-editing macros can produce sizable increases in productivity.

In conclusion, we believe that the combination of powerful automation and electronic resources, plus hybridization of methods will gradually raise the glass ceiling. The history of the state of the art of MT has been one of evolutionary improvement rather than miraculous breakthroughs. The recent introduction of statistical methods and other automated methods has not so much raised the glass ceiling, as they have added to our bag of tricks with which to attack the well-known problems. Concerning fully automatic MT itself, we take the view that MT for assimilation is an unqualified success for general text. The explosive growth of available information on the Internet, combined with corporate multilingual e-mail, workflow, and intranet technologies, as well as the emergence of cross-language retrieval engines are the guarantee for an exiting future in the language software industry.

So maybe the question is not so much if we are optimists or pessimists about MT and automation, but for what purpose are we developing? Besides, the fact that we are all present at this conference automatically qualifies us as optimists. The pessimists have already gone on to other challenges, and will miss all the fun.

Panel Position On the Limits of Automation

Yorick Wilks

University of Sheffield

A skeptic might infer from the IBM statistical MT program, and its now well-understood limitations, that that alone showed the limits of automation in MT in a dramatic manner. On that account, IBMs methodology was to MT only from automated data and the fact the system seemed to "top out" somewhere around 50% of sentences correctly translated showed that was all automated resources could achieve, confirmed by the fact that the IBM team then turned to gathering more conventional linguistic resources such as lexicons and grammars.

But that account, not one any researcher has argued in print of course, is far too simplistic: the later move was not away from automated resources but directed towards the gathering of a different sort or level of automated resources, namely grammars and lexicons. This point, if I am right, confirms that automation is not, as some would assume, connected in any way with any particular level or type of linguistic resources but is simply an acquisition methodology. If there is only very little (there is some!) data on human pragmatics acquired automatically, it is perhaps only because few people have looked for it, not because it is

inherently absurd to look. This point may be obvious to many but it is not yet uniformly accepted, I am certain.

What also muddles discussion of this issue is that automatic acquisition is often from resources that are, in whole or part, much more than human language production: some, like MRDs, are explicitly intended as resources. This fact does not shift the practical issue of resource acquisition for NLP—one should get what one can, wherever one can—but it does shift the issue if, anywhere in the back of ones mind is a Chomsky-style analogy between acquisition for NLP and first language learning. First language learners do *not* have access to MRDs or bilingual parallel text, so acquisition from them cannot be in any way related to arguments about how humans acquire a first language *and what existing apparatus, if any, that acquisition requires.*

This is, for me, the nub of the practical matter, and why I remain pro-automation as far as it can go but do not believe full automation, unsupervised, and from raw unprimed text or dialogue, can provide all we need, no matter what the outcome of outstanding issues in learnability theory. So, I take a pressing research question, as do many others, to be: how do we tune a lexicon to a new corpus, automatically if possible. But note that that formulation assumes we already have a human-created resource, i.e., the lexicon we started with. It is thus:

I: structure1 + corpus —> structure2

which indicates that an earlier state of the structure itself plays a role in the acquisition, of which structure2 is then a proper extension (capturing new concepts, senses etc). This is a different model from the wholly automatic model of lexicon acquisition often used in, say, TIPSTER related work, which can be written:

II: corpus —> structure

This latter case is one which does not update or 'tune' an existing lexicon but derives one directly and automatically from a corpus. I am arguing the essential role of a representational structure in this process, and hence *for* the first process where possible—which we may also take to involve some essential human intervention as well.

The assumption here is that we cannot understand the nature of the representation of meaning in lexicons, or elsewhere, unless we can see how to extend lexicons in the presence of incoming data that does not fit the lexicon we start with. Extension of representation is part of an adequate theory of representation.

However, what are we to do in the pressing practical situation of wholly unknown languages, ones perhaps with electronic texts available, but no lexicons: in that case, only mode II is possible initially. But, almost certainly, on the above assumptions, human informants will also be necessary to move from mode I to mode II, so as to have a lexicon to tune. It is, at the moment just a hope that processes like II, with human assistance, will provide lexicons for tuning processes like I.

Even so, I still believe we ought to act as if all possible automation should be used wherever possible, and that our encyclopaedist/lexicographic tendencies should be repressed, as leading to vast resource creation that is never properly made use of because the purpose and the real task ahead cannot not be fully foreseen. Automation cuts down the drudgery and keeps us sane and focussed, so let us limit human intervention as far as we can.