Spoken Language Translation

Introduction

Steven Krauwer
Doug Arnold
Walter Kasper
Manny Rayner
Harold Somers

Some 15 years ago, when Machine Translation had become fashionable again in Europe, few people would be prepared to consider seriously embarking upon spoken language translation (SLT). After all, where neither machine translation of written text, nor speech understanding or speech production had led to any significant results yet, it seemed clear that putting three not even halfway understood systems together would be premature, and bound to fail.

Since then, the world has changed. If we look at the papers contained in the proceedings of this workshop we can clearly see that many researchers, both in academia and in industry, have taken up the challenge to build systems capable of translating spoken language. Does that mean that most of the problems involved in speech–to–text, text–to–text translation, and text–to–speech have been solved? Or should we rather conclude that all these courageous people are heading for another traumatic experience, just as we have seen happen in the sixties and, to a lesser extent, in the eighties.

The answer to the first question is probably: No – although we have made a tremendous progress, both from a scientific and from a technological point of view, many of the fundamental problems in MT and in speech understanding remain unsolved. Yet we are convinced that the bleak scenario we mentioned as the alternative does not apply either.

There are a few reasons why we feel confident that a certain degree of optimism is justified here. First of all, it is clear that on the whole people's expectations of what MT will do for them are changing. Where in the past the ultimate goal of MT seemed to be to provide a perfect, but cheaper and faster alternative to the human translator, there is now a clear shift from the ideal of fully automated high quality translation of unrestricted texts to the more practical problem of overcoming the language barriers we encounter in various situations. This shift of focus allows us to partition the problem we address into a series of smaller ones, the solution to which may be within our reach. In other words, instead of trying to win the war against an enemy we are not even sure we can see, we have decided to engage into a series of battles we can be confident of winning.

This applies both to spoken and written language translation. If we look at spoken communication between human beings with different native languages, very often the main success criterion for this communication is not whether or not the individual sentences produced by the participants have been expressed or understood without errors (which will rarely be the case), but rather whether the intended goal of the communication has been attained (hotel room reservation, airline information, etc). This observation is extremely important when we try to set our goals for spoken translation systems. Once we have realized that communication takes place in a specific context, with a specific goal, and have accepted that sentence–by–sentence linguistically correct translation is not a necessary condition for successful multilingual communication, we can start exploiting the full potential of spoken dialogues in human–human and human–machine interaction: the basic structure of dialogues, the ways to control dialogue flow, the possibility for repair.

To summarize, although many of the fundamental problems of MT and speech have not been solved, the movement towards more specialized systems, the redefinition of the notion of success, and the potential of dialogues, taken together, give us reason to believe that we will see many successful spoken translation systems in the near future, and we hope that this workshop will contribute to this.

In the rest of the introduction we will introduce very briefly the topics of the four sessions of the workshop.

In the proceedings one will also find three 'poster papers'. Although the workshop session itself did not leave space for poster presentations, we felt that it was important to dedicate a small section of the proceedings to short poster papers, where researchers can communicate to others what they are doing, so that people who are interested in the same or related research topics know where to go.

Exploiting and Exploring Dialogue Structure

SLT is the latest frontier for MT research – perhaps the last frontier. A term sometimes seen used is "Machine Interpreting", but it seems that this might apply to only one aspect of SLT, implying some activity similar to that of human interpreters, i.e. simultaneous or consecutive translation of spoken language, often in the context of a meeting or someone addressing a group of people. Notice that such speech may or may not be wholly spontaneous. This contrasts with the type of SLT which is the theme of this first session, and indeed more predominantly influences SLT research so far, namely Dialogue Translation. Let us note in passing a third type of SLT, "Message passing", for example so-called "voice-mail", or real-time messages between emergency or security services across linguistic borders (e.g. the Channel Tunnel).

Within the subdomain of Dialogue Translation, we can make some further relevant distinctions, all of which will impinge on the design of the MT system: telephonic vs. face-to-face dialogue, co-operative vs. adversarial (Kay *et al* 1994:175f), involving completely or partially monolingual speakers, with or without system–user "meta-dialogue" (Somers *et al* 1990), and so on.

Dialogue MT introduces interesting problems beyond the already difficult issues of integrating speech processing with translation. As has clearly been recognised in the three papers which make up this opening session, identifying the special pragmatic features of spoken dialogue which distinguish such "texts" from the type of input that a traditional MT system might deal with is a crucial part of the problem. Traditionally (e.g. Hutchins & Somers 1992:92), the incorporation of contextual knowledge into an MT system was just dismissed as impractical, or at best, uneconomical. In a dialogue system, such an approach is unthinkable.

Manfred Stede and Birte Schmitz initiate the proceedings with a close look at "discourse particles", the little words which can carry so much meaning, especially in terms of the overall dialogue structure. An additional problem is that many of these particles are ambiguous in that they also have an interpretation not related to discourse structure.

Jae-won Lee *et al* look at words whose translation is particularly dependent on the context, a problem which is exacerbated in a language-pair such as Korean–English. Their approach is to apply a statistical model of dialogue structure based on trigrams of speech acts.

Keiko Horiguchi discusses meaningful "errors" in speech which convey contextual meaning or the speaker's attitude, and then focuses on the translation of discourse particles from Japanese into English. The approach adopted here is an analogical framework using a Cascaded Noisy Channel Model.

Dealing with Differences

Although translation of written and spoken language have much in common, there is no evading the fact that text and speech are in some ways fundamentally different modalities. The impermanent nature of vocal communication makes speech an intrinsically more unreliable medium; conversely, a spoken utterance contains information that is only residually present in its text version, such as prosody, tone and accent.

The three papers in this section explore some aspects of the SLT problem which highlight the differences between translation of written and spoken text. Yumi Wakita *et al* describe a method which attempts to extract the parts of a spoken utterance which have been reliably recognized, ignoring those which represent probable recognition errors. They present results indicating that their method has an appreciable effect on the performance of a Japanese–English speech translation system.

Keiko Horiguchi and Alexander Franz describe another piece of work aimed at counteracting the problems involved in taking translation input from a speech recognizer. They present an example-based hybrid approach containing aspects of both corpus-based and rule-based styles of translation architecture; this move towards hybrid architectures seems to represent a strong tendency in current work within the field of SLT

Finally, Pascale Fung *et al* present a paper focussed on the problems which a speech recognizer has to contend with in a multilingual environment, where people typically speak using a variety of languages and accents. The paper describes initial experiments which investigate the parameters of the problem, and in particular explores the possibility of constructing recognizers capable of recognizing multilingual input.

Towards Efficiency

The papers in this section address problems of efficiency in two senses. On the one hand SLT has to meet specific requirements of efficiency and robustness in *processing*, because

- speech recognition is imperfect and the input is often not a string but a lattice of word hypotheses representing a set of possible utterances
- spoken utterances are often linguistically not well-formed
- the translation must be available nearly simultaneously with the utterance
- the quality of the translation must be sufficiently high as in most applications post-editing is not possible. The approach of Frederking *et al* differs in this respect as it allows for user interaction to improve the translation.

To solve these problems finite state transducer technologies are often employed and investigated. The papers by Alshawi *et al* and Amengual *et al* discuss different approaches along these lines. Both attempt to gain additional efficiency by a tight integration of analysis and transfer instead of assuming two different processing stages.

Another efficiency problem is that of acquiring the knowledge for building such a system. SLT systems are often heavily restricted to specific domains and in their vocabulary. This raises the question how such systems can be adapted to new domains and vocabulary. Therefore corpus-based statistical methods for language modelling and automatic acquisition are of special interest for SLT as addressed in in Amengual *et al* and Frederking *et al.*

Methodological Issues

The common theme for the final three papers in the workshop is an emphasis on methodology and architecture. In the first two, the focus is on the methodology required when one moves from one application domain to another. In the case of the first paper, it is the move from smaller domains to larger more inclusive domains, in the case of the second, it is the move "across" from one domain to a distinct and separate domain of similar size. Both papers explore the sorts of approach and architecture that the different sorts of move require. Given that the state of the art in Speech Translation is such that realistic applications are restricted to particular domains, this sort of study is clearly of general importance. The third paper, by Mark Seligman, takes a broader methodological and architectural perspective, and identifies six issues of importance to the field as a whole.

The first paper in the section, Lavie *et al*, focuses on the issues that arise when one transfers from a relatively narrow domain (in this case, Appointment Scheduling dialogues) to a broader domain (Travel Planning dialogues). The paper describes some preliminary results of making this transfer for the JANUS system, and some modifications that may be required. Differences between smaller and larger domains include a higher out-of-vocabulary rate, a higher rate of ambiguity, and generally the existence of a much wider range of expressions and expressive devices in dialogues which make the 'semantic grammar' approach — which worked well in the narrower domain — problematic. Lavie *et al*'s suggestion is that this problem and the problems that arise from increased ambiguity can both be overcome if the larger domain can be factored into a number of sub-domains, each of which can be given its own semantic grammar. Such sub-grammars should be far less ambiguous than a grammar for the whole domain would be, if parsing proceeds with separate sub-grammars in parallel, which should also yield benefits in terms of processing speed.

In the second paper in this section (by Carter *et al*), the main issue is not how one can broaden or enlarge the domain of a system, but how one can move from one domain to a distinct, potentially unrelated, domain of similar size (or even to a distinct language pair, which can raise similar issues — this should be clear if one compares moving between pairs of very similar languages which may share a great deal of grammar and vocabulary with moving between very different domains which share very few features of grammar and vocabulary). In other words, the focus is on the problems of customizing systems for new domains and languages. Carter *et al* argue that the characteristics of the Core Language Engine — the language processing component of the system they are describing (SLT) — facilitate this customization. In particular, they suggest that the use of a general-purpose linguistic rule component, and a transfer architecture, in combination with statistical information derived from supervised training on corpora make most of the SLT system portable across domains, and even languages, and the remaining, non-portable, parts of the system are such that they require relatively little expert knowledge. This conclusion is interestingly at variance with that of Lavie *et al* in the previous paper, who argue for an interlingual approach to translation and the use of domain-specific semantic grammars.

The final contribution in this section is Mark Seligman's, which takes a personal perspective in identifying six areas of SLT research as particularly interesting. (1) He argues the need for interactive disambiguation (a view that the authors of the other papers in this section would probably reject), and (2) for a particular kind of system architecture (a variant of the blackboard architecture incorporating a supervisory coordinator program) which may also be controversial. (3) The third issue he addresses is that of how Speech Recognition and MT techniques should be integrated — in particular, whether a single set of techniques can or should be used to cover both tasks, e.g. parsing to the level of phones. Seligman suggests that this is promising, though there are technical problems. (4) Seligman's fourth issue is how far natural pauses can be used in segmenting utterances, and how far analysis and translation can proceed on the basis of such segmentation. (5) The fifth issue recognizes the importance of Speech Act identification in dialogue translation, and considers how a defensible and usable classification may be found. (6) Finally, there is the question of how one can restrict the range of candidate lexical items that have to be considered at each point in processing, and how candidates can be weighted appropriately. Seligman observes that accepting the importance of these issues suggests a particular architecture for an experimental SLT system which differs from systems described in other contributions in significant ways.

Concluding Remarks

As we welcome delegates to what we believe is the first major open meeting in Europe devoted entirely to SLT, but surely not the last, we signal yet another important milestone in the history of Machine Translation. Just fifty years since Warren Weaver, in his letter to Norbert Wiener (later reproduced in his famous memorandum), first expressed realistic hopes for "mechanical translation" (see Hutchins, in press), we find ourselves realistically discussing the possibility of using computers to translate the *spoken* word. Dismissed not so long ago as an impossible dream, the contributions to this workshop demonstrate that, while still perhaps something of a dream, it is far from impossible. As the world of MT looks for new directions, SLT offers a wide range of new challenges. This new focus will be reflected in a Special Issue of the journal *Machine Translation* devoted to SLT, for which a call for papers will be issued soon; and already we can see, in other MT-related conferences and publications, a clear inclination towards this problem area. Let us hope that in years to come, the Workshop on Spoken Language Translation at the 1997 ACL/EACL meeting in Madrid is seen as an important and memorable event in the development of SLT techniques.

References

Hutchins, W. John and Harold L. Somers. 1992. *An Introduction to Machine Translation*, London: Academic Press.

Hutchins, John. in press. From first conception to first demonstration: A chronology of the nascent years of machine translation, 1947–1954. To appear in *Machine Translation*, **12** (1997).

Kay, Martin, Jean Mark Gawron and Peter Norvig. 1994. *Verbmobil: A Translation System for Face-to-Face Dialog*, CSLI Lecture Notes No. 33, Stanford, CA: Center for the Study of Language and Information.

Somers, Harold L., Jun-ichi Tsujii and Danny Jones. 1990. Machine Translation without a source text. In *COLING-90: Papers presented to the 13th International Conference on Computational Linguistics*, Helsinki, Vol.3, 271–276.

Programme and Organizing Committee

Steven Krauwer	ELSNET and Utrecht Institute of Linguistics OTS, Utrecht University *(chair)* *steven.krauwer@let.ruu.nl*
Doug Arnold	Department of Language and Linguistics, University of Essex, Colchester *doug@essex.ac.uk*
Walter Kasper	DFKI, Saarbrücken *kasper@dfki.uni-sb.de*
Manny Rayner	SRI International, Cambridge *manny@cam.sri.com*
Harold Somers	Department of Language Engineering, UMIST, Manchester *harold@ccl.umist.ac.uk*