Worldwide Ongoing Activities On Multilingual Speech to Speech Translation

Gianni Lazzari , *Alex Waibel**, C. Zong****

*ITC-irst Sensory Interactive Systems Division Trento Italy lazzari@itc.it
** CMU Interactive Lab Pittsburgh US aw@cmu.edu
*** National Laboratory of Pattern Recognition, Beijing, China. cqzong@nlpr.ia.ac.cn

Abstract

This paper presents an overview of worldwide going on activities on Speech-to-Speech Translation. After a short introduction of the field, including the major projects and milestones, activities and projects going on in Asia, Europe and US are presented and described.

1. Introduction

Speech Translation (ST) or Speech-to-Speech translation (SST) research topic represents a quite recent research area in the Human Language Technologies arena. Some important dates need to be mentioned: 1986, in Japan the ATR's project on SST starts, 1992 birth of C-STAR consortium <http://www.c-star.org>, in 1993 the German government funds the national project Verbmobil <http://verbmobil.dfki.de/overview-us.html>, in 2000 DARPA launched Tides < http://www.darpa.mil/ipto/Programs/ tides/index.htm >, in 2004 the European Commission funded TC-STAR <http://tc-star.org>. Looking back to these twenty years, the ST research has followed the typical path of a new area: starting from a usage scenario (interpreting telephony, e-commerce, face to face) during the first 10 years and arriving to the definition of a research agenda (objectives, tasks to be afforded, methods and evaluation). A key role has been played by C-STAR consortium in demonstrating the feasibility of the SST technologies through two main worldwide life demonstrations in 1995 and 1999. Verbmobil also played an important role in the community, trying to afford both a feasibility demonstration of a face to face communication mediated by an automatic translator and on the same time explore and evaluate different approaches, semantic transfer, dialogue based act, example based and statistical. TIDES has been an innovative program focusing on the Trans lingual Information Detection, Extraction and Summarization.(see below). It is quite worth noting that the move from demonstration based activities to evaluation based has been the path followed by the ASR community. We expect that SST will show a similar development as in speech recognition thirty years ago. Program like TIDES, the evaluation campaigns carried on in C-STAR III <http://www.slt.atr.co.jp/ IWSLT2004/> and the new European integrated project TC-STAR, which will be based on yearly evaluation campaigns demonstrate that this research sector can face a new maturity phase, which is characterized by:

- a diffuse consensus, shared also by the Machine Translation (MT) community, on data-driven approach, where statistical methods seems outperform and where the approach can be summarized as the combination of Language Modeling and Statistical Decision Theory;
- the definition of a framework of research development based on the paradigm of competition and cooperation. This means that a community will focus mainly on common tasks, using common data and competing on the methods and algorithms. This process will fasten the research development even more that the one experienced within the ASR community if competition will also be supported by a strong cooperation avoiding duplication of code writing through the adoption of an open source methodology. (An interesting experience is carried on by the bioinformatics community);
- Worldwide research cooperation, involving all the major sites in the world participating to this grand challenge. A critical mass of researchers need to be involved and a huge amount of multilingual data are necessary. There will be no advance in the field if this critical mass will not be reached.

In what follows the major projects and activities in Asia, US and Europe will be presented and discussed..

2. ST activities in ASIA

In Asia, SST has been an attractive research topic for many years. The Interpreting Telephony Research Laboratories (ITL) of Advanced Telecommunications Research Institute International (ATR) <http://www.atr.jp> of Japan was established in 1986, which is the first laboratory working with the speech-to-speech translation in Asia, and is one of the sponsors of C-STAR organization. In the first phase of the history of speech translation research theme in ATR (from 1986 to 1992), the target was focused on probing into the feasibility of speech translation technology. The first international joint experiment of interpreting telephony between Japan, USA, and Germany was successfully conducted in January of 1989, in which the inputs was restricted as the grammatical correct expressions and clear utterances.

In the second phase, from 1993 to 1999, the research target was moved to study the applicability of natural dialogues, especially to cope with the input with ellipses, fragmental expressions, and ungrammatical sentences in the natural spontaneous speech. The experimental or prototype systems, which translate spoken Japanese into Korean, German, English and Chinese, was developed in 1997, and the ATR-MATRIX Japanese-English bi-directional speech translation system was developed in 1998. The second international joint experiment of multilingual speech translation was successfully conducted in the July of 1998. And also, the wearable speech translation system and Japanese-English bi-directional speech translation system operating on a notebook PC was successfully developed in the July of 1998. In October of 1999, the ATR-MATRIX was successfully extended to a speech translation system that is easily accessible for users with a cellular phone.

From 2000, the research on SST in ATR has been the third phase, whose target is to study the possibility in the real world applications and the approaches to deal with the wide range of expressions, including the ungrammatical sentences and the utterances under various environments are specially addressed.

In China, the first project on SST supported by China government was an important national science foundation (NSF, <http://www.nsfc.gov.cn/>) project under grant number 69835030, which was taken charge by the Institute of Automation, Chinese Academy of Sciences (CAS-IA, < http://www.ia.ac.cn/ >), from January of 1999 to December of 2002 [1,2]. In this project some key issues of SST and the paradigm to realize a high performance multilingual spoken language translation (SLT) system had been investigated.

In 2002, the China High-Technical Program (863 Program) < http://www.863.org.cn > funded an important big project named Research on Key Technology and Development of Demonstration System for Multilingual Intelligent Information Service Oriented to Olympic Games (the grant number is 2002AA117010). This is the so-called "Digital Olympic" project. The project principals are from Beijing Capital Information (Cap Info) Co. Ltd. <http://www.capinfo.com.cn> and CAS-IA. This project is focused on the investigation of the technical key issues to develop the practical intelligent information service system for 2008 Beijing Olympic Games. The research topics include many aspects on the language and speech technology, including the machine translation oriented to the WebPages translation, SST, information retrieval, and information service based on mobile terminals and so on. Considering the importance and the grand significance of the project, Beijing Science and Technology Committee also supports some funding. This makes the "Digital Olympic" project so arrestive and attractive. The research groups led by Alex Waibel in the University of Carnegie Mellon, USA, and the University of Karlsuhe, Germany, fortunately join this project as an only partner outside China to work with the SST sub-task in cooperation with CAS-IA. For the multilingual SST, the Chinese side, CAS-IA, takes charge of the Chinese corpus collection and analyzing, Chinese speech recognition, Chinese input-to-interchangeable semantic form (IF) parsing, IF-based generation, and also the Chinese speech synthesis as well. The first version of the joint SST system was successfully demonstrated in Beijing exposition of sciences and technologies from 22nd to 26th May 2004. In the coming July, the system will be also demonstrated in Barcelona forum of world culture. CAS-IA also has made much achievement on the PDA-based SST technology.

Oriented to the development of practical SST system for Olympic Games, ATR-SLT, ETRI of Korea, and CAS-IA jointly signed the CJK (Chinense-Japansese-Korean) cooperation agreement. The purpose of the agreement is to joint research and develop an SST system, which converts spoken expressions (voice) related to travel conversation between the three languages (the Japanese language, the Chinese language, the Korean language).

ETRI of Korea has been performing the research on speech translation as a part of the government project, titled by "Language information processing technology development." This project focuses on how to use the speech translation technology in real life. This is because the feasibility of the technology had been shown in the previous demonstrations, such as ETRI-KT-KDD demonstration on hotel reservation domain in 1995 and C-STAR II demonstration on travel planning domain in 1999, but it is not commercialized yet. With this regard, ETRI has targeted on the mobile phone speech recognition, translation of simple but crucial expressions for traveler, and dialogue style speech synthesis. The project also includes the development of the technologies of Korean broadcasting news speech recognition with vocabulary of 65,000 words and Korean-to-Chinese broadcasting news text translation with a 200,000 word lexicon. By combining these technologies, it is possible to attack monologue style speech translation, like lectures or presentations as well as the broadcasting news speech.

3. ST activities in Europe

Language is a topic of major importance for the construction of the European Union, which is of economical, cultural, political and social nature. The effort to address such crucial issue appears to be too large for the Commission alone, and it seems to be necessary to have this effort shared between the Commission (which has the duty to ensure a good communication with the member states, and among the member states), and the member states which have to preserve and promote their language(s), and through their language(s), their culture. For this reason the Commission, in the V and VI frameworks, identified HLT as strategic objective. Presently there are three active projects: LC-STAR <www.lc-star.com>, PF-STAR <www.pfstar.itc.it> TC-STAR <www.tc-star.org>. Previously the most important projects carried on in Europe have been, Verbmobil I and II funded by the German government and EU-TRANS and NESPOLE! funded by the European Commission.

- **LC-STAR**, launched officially in 1st of February 2002, is a project focusing on creating language resources (LR) for transferring Speech-to-Speech Translation (SST) components and thus improving human-to-human and man-machine communication in multilingual environments. The goal of the project is to create lexica for 13 languages and text corpora for 3 languages and to create a demonstrator translating within 3 languages. The work is performed by two parallel running tracks of which Track I (duration 2 years) will concentrate on specification and creation of large word lists and lexica while the Track II (duration 3 years) will concentrate among others on investigation, specification and creation of LR needed in speech centered translation technologies and also building a demonstrator for speech to speech translation.

The lexica will cover 13 target languages. Each lexicon will have at least 100 000 entries consisting of 50 000 common words, 45 000 proper names and 5 000 application specific words. The corresponding word lists are extracted from large text corpora containing at least 10 million of words. The language covered are Italian, Greek, Russian, Turkish, Spanish, Catalan, German, Classical Arabic, Hebrew, US-English, Finnish, Mandarin Chinese, Slovenian. In the second phase aligned bilingual text corpora and monolingual lexica with morpho-syntactic entries will be produced.

The size of the text corpora and the kind of morpho-syntactic information needed will be evaluated with respect to speech translation quality. Based on the evaluation results lexica for 8 languages and bilingual text corpora for 3 languages will be specified and created. These LR will cover a tourist domain, which is also chosen for the demonstrator showing the language transfer within 3 languages. The following language pairs are covered for aligned monolingual lexica and bilingual text corpora: Catalan / US-English, Spanish / Catalan, Spanish / US-English.

- **PF-STAR** addresses the challenging goals of

providing advanced technological baselines, (comparative) evaluations and assessment of prospects for three key technological areas: speech-to-speech translation (STST), the detection and

expressions of emotional states, and core speech technologies for children. The project builds on years of research already under way under various national and international research projects, most notably NESPOLE!, C-STAR, Verbmobil, SmartKom. The languages considered are English, German, Italian, and Spanish.

PF-STAR results consist of technological baselines, assessed with respect to both observed performances and future prospects, and linguistic databases. No demonstrations or showcases are planned and no particular scenario will be addressed.

Partners of the project are: ITC-IRST, RWTH, UKA, CNR ISTC-SPFD, EURLN, KTH, and UB. The project started in September 2002 and will last 2 years.

- Finally on April first **TC-STAR** project started.

TC-STAR is a integrated project with 11 Million Euros grant and 12 partners: ITC-IRST, RWTH, LIMSI, UPC, UKA, IBM, Siemens Germany and France, Nokia, Sony Stuttgart ELDA and SPEX. TC-STAR is envisioned as a long term effort focused on advanced research in all core technologies for speech to speech translation (SST): speech recognition, speech translation and speech synthesis.

The objectives of the project are extremely ambitious: making a breakthrough in SST research to significantly reduce the gap between human and machine performance. The focus will be on the development of new, possibly revolutionary, algorithms and methods, integrating relevant human knowledge, which is available at translation time into a data-driven framework. Examples of such new approaches are the integration of linguistic knowledge in the statistical approach of spoken language translation, the statistical modeling of pronunciation of unconstrained conversational speech in automatic speech recognition, and new acoustic and prosodic models for generating expressive speech in synthesis.

TC-STAR is planned for a duration of six years, which is the time needed for exploring and evaluating new approaches to SST, and for creating the infrastructure needed for accelerating the rate of progress in the field. The project has been divided in two phases of three years length. The first three years of the project's work-plan has been granted and will target a selection of unconstrained conversational speech domains – i.e. broadcast news and speeches – and a few languages relevant for Europe's society and economy: native and non native European English, European Spanish and Chinese. The second three years, will target more complex unconstrained conversational speech domains – i.e. meetings and social conversations – adding to the previous languages other relevant European languages. This second phase will give rise to a new proposal for funding to be evaluated in a later FP6 competitive call. Key actions for reaching the objectives and affording these grand challenges will be:

· the implementation of an evaluation infrastructure based on competitive evaluation, in order to achieve the desired breakthroughs

· the creation of a technological infrastructure aimed at fostering the effective delivery and evaluation of scientific results

· the support of knowledge dissemination of scientific results within the consortium and the research community.

4. ST activities in US

Even though speech translation research in the United States has begun already in the early 90's, recognition of its importance has developed only gradually. This has been caused in part by the dominance of English as today's lingua franca, the language by which international business, science and commerce is carried out worldwide. As such multilingual speech and language processing have frequently been seen as a secondary priority in science and commercial development. Geopolitical events, US involvement globally, international trade and business, and last not least, changes in demographics have now changed this perception dramatically. While it is generally more difficult for English native speakers to learn other languages (for lack of opportunity in a large mono-lingual society), the need for foreign language capability has become apparent. The need for better access to foreign language documents and multimedia information (that is not necessarily in English) has been recognized by business and government alike. Similarly, communication with non-English speakers locally in foreign field situations was identified as a critical need for military and humanitarian relief personnel. The need, however, does not only exist for assignments abroad, but within the United States: First responder services, medical services, fire fighters, and many more require translation support on site, to respond more effectively to emergencies and human needs in a country, whose linguistic make-up is rapidly changing and becoming more diverse.

In view of these changes, several new research initiatives have been launched in the US, elevating Speech Translation to one of the key objectives in speech and language technologies. Three current on-going projects aim to address these language needs:

- **DARPA TIDES** – TIDES is a large language

technology effort aiming to produce high quality Tran lingual Information Detection, Extraction and Summarization (TIDES) of (multilingual) texts. Machine Translation (MT) is one of the greatest challenges in TIDES, and the new push promises to produce dramatic new advances after two decades of comparatively limited work in MT. TIDES does not include speech, but basic MT technologies are developed that influence and inspire work in the speech translation area as well. The main objective for TIDES is assimilation of multilingual content from text. Key languages are English, Chinese, Arabic, and additional ("surprise") language experiments are carried out to explore portability of the technology. Research laboratories involved in MT under TIDES include: CMU, IBM, ISI, JHU, RWTH.

- **DARPA Babylon** – The ***BABYLON*** project and its follow-in project CASTE, differ in their mission from TIDES, in that they attempt to provide *interactive* cross-lingual support in field situations. Here cross-lingual dialog by voice between and English and non-English language speaker is the key objective. Target scenarios include refugee processing, medical triage, force protection. Languages include: Arabic, Chinese, Pashtu, Farsi, and Thai. The systems under development are all domain limited, the two-way systems use an Interlingua based approach. Some attempt speech translation only in one direction (one-way), both directions (two-way), or a so-called one-plus-one way, where speech translation in one direction is complemented by phrase and concept spotting on the foreign language response. The Babylon systems are also ported to small hand-held PDA-devices to permit user tests in mobile field situations. Implementations both for commercial off-the-shelf platforms as well as a ruggedized PDA specially developed for field use by Marine Acoustics (the "Phraselator") have been demonstrated. Laboratories participating in Babylon speech translation development are: BBN, CMU, IBM, ISI/HRL, SRI.
- **NSF STR-DUST** – STR-DUST is an NSF supported 5 year initiative that attempts to go beyond domain-limited speech translation and investigate open-domain speech translation. The project is carried out at CMU and languages are Chinese, Arabic and English. The project is in a basic research phase, exploring techniques to deal with the double challenge of unlimited topical/semantic scope of the domain (and vocabulary) at the same time as the disfluent, syntactically ill-formed form of spoken language.

In addition to these major government sponsored research initiatives, several commercial developments and collaborative activities exist. Several companies (Ectaco, Marine Acoustics, among others), now develop and commercialize small handheld voice operated phrase-books for tourists and military personnel. These devices do not provide fully unconstrained two-way speech translation, but permit voice input to lists of typical phrases for look-up and voice synthesis. Short of full dialog translation, they are a pragmatic and effective first generation tool toward bridging the linguistic divide.

Among academic and research laboratories, several more informal and/or unfunded collaborations are instrumental in moving the state of the field forward. Notably, the worldwide C-STAR consortium engages several key US partners and affiliates. Among them: Carnegie Mellon (as partner) in collaboration with University of Karlsruhe (Germany); ATT, IBM, Lincoln Laboratories, MIT (affiliates). C-STAR promotes joint development of prototype systems as well as joint evaluations of speech translation performance. Other collaborations include bilateral collaborations, such as CMU's participation in the Chinese Digital Olympics effort, Barcelona's Forum 2004, the EC/NSF Nespole! program, and others. Strategies being researched focus on four major methods: Interlingua Based Translation (IBT) via semantic grammars, IBT via statistical learning methods, and Example Based Translation (EBMT) and Statistical Translation (SMT). Successful systems have been demonstrated with all four methods. Evaluations are being carried out by individual laboratories as well as government agencies at the component level, end-to-end and by user-studies.

5. Conclusions

In this paper worldwide ongoing activities and projects on speech to speech translation going on in Asia, Europe and US have been presented.

6. Acknowledgements

The authors would like to thank Dr. Hiromi Nakaiwa, ATR-SLT of Japan, and Dr. Jun Park, ETRI of Korea, for their beneficial references.

7. References

[1] A. Waibel, A. Jain, A. McNair, H. Saito, A. Hauptmann, J. Tebelskis "JANUS: A Speech-to-Speech Translation System Using Connectionist and Symbolic Processing Strategies". In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Toronto, May, 1991.

[2] Franz Joseph Och, Hermann Ney: A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics* 29(1): 19-51 (2003)

[3] Yamamoto, Seiichi. 2000. Basic Research on Speech Translation Technologies. In *ATR Journal*, Vol. 3. pp. 27-29.

[4] Zong, Chengqing, Taiyi Huang and Bo Xu. 2000. Design and Implementation of a Chinese-to-English Spoken Language Translation System. In *Proceedings of the International Symposium of Chinese Spoken Language Processing (ISCSLP)*, October 13-15, 2000. Beijing. Pages 367-370.

[5] Stephan Vogel, Alicia Tribble. Improving Statistical Machine Translation for a Speech-to-Speech Translation Task. In *Proceedings of ICSLP-2002 Workshop on Speech-to-Speech Translation*. Denver, CO. September 2002.

[6] Lazzari G. Spoken Translation: Challenges and Opportunities. In B. Yuan, T. Huang, X. Tang (eds.) *Proceedings of 6th International Conference on Spoken Language Processing*, Beijing, China, October 16-20, 2000, vol. IV, pp. 430-435

[7] Rudnicky, A. I., Polifroni, Thayer, E H., and Brennan, R. A. "Interactive problem solving with speech", *J. Acoust. Soc. Amer., Vol. 84, 1988, p S213(A)*