Multilingual Mobile-Phone Translation Services for World Travelers

Michael Paul, Hideo Okuma, Hirofumi Yamamoto, Eiichiro Sumita,
Shigeki Matsuda, Tohru Shimizu, Satoshi Nakamura
† NICT Spoken Language Communication Group
‡ ATR Spoken Language Communication Research Labs
Hikaridai 2-2-2, Keihanna Science City, 619-0288 Kyoto, Japan
Michael.Paul@nict.go.jp

Abstract

This demonstration introduces two new multilingual translation services for mobile phones. The first translation service provides state-of-the-art text-to-text translations of Japanese as well as English conversational spoken language in the travel domain into 17 languages using statistical machine translation technologies trained automatically from a large-scale multilingual corpus. The second demonstration is a speech translation service between Japanese and English for real environments. It is based on distributed speech recognition with noise suppression. Flexible interfaces between internal and external speech translation resources ease the portability of the system to other languages and enable real-time location-free communication world-wide.

1 Introduction

Spoken language translation technologies attempt to bridge the language barriers between people with different native languages who each want to engage in conversation by using their mother-tongue. The importance of these technologies is increasing due to increases in the number of opportunities for cross-language communication in face-to-face conversation, especially in the domain of tourism.

We demonstrate two multilingual translation services for mobile phones that are built on corpus-based speech recognition and translation technologies. These services enable smooth and location-free communication in real environments covering the major languages of most nations (see Figure 1).

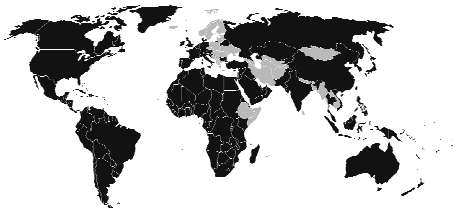

Figure 1: Global Language Coverage

The first multilingual translation service described in this paper is a text-to-text translation service that enables users to translate Japanese and English conversational spoken language sentences in the travel domain into 17 other languages. The system's core components consist of a multilingual, sentence-aligned spoken language corpus covering 18 of the major world languages and state-of-the-art statistical machine translation (SMT) engines that are trained automatically from this corpus covering 306 (=18x17) translation directions. A graphical user-interface (GUI) allows 24x7 world-wide access to the translation service (see Section 2).

The second multilingual translation service is an extension of the text-based translation service that additionally provides speech recognition capabilities. This is the first commercial speech translation service in the world. The system is based on distributed speech recognition and operates as follows: (1) front-end processing (noise suppression, feature extraction, and feature parameter compression) is carried out on the mobile phone, (2) back-end processing (recognition, translation) is done on a server and (3) translation results are sent back and displayed on the mobile phone (see Section 3).

2 Multilingual Text Translation Service (MTTS)

The multilingual text translation service for mobile phones can be accessed via 'http://atr-langue.jp/smlt' or by using the QR code in Figure 2 that also illustrates the graphical user interface of

Coling 2008: Companion volume – Posters and Demonstrations, pages 165–168
Manchester, August 2008

Figure 2: QR Code and GUI of MTTS

the translation service. Two different modes are distinguished: (1) the *multilingual mode* where the input is translated into all 17 languages simultaneously and the translation results are displayed side-by-side and (2) the *bilingual mode* where a single language out of 17 languages can be selected as the target language of the Japanese or English input text translation. The *bilingual mode* also features *back-translation* functionality, i.e., a reverse translation of the generated translation output into the source language, that enables immediate feedback on the quality of the translation output. In order to solve font problems of mobile phones, the translated sentences are rendered on the server side and an image is sent and displayed in the mobile phone.

2.1 Multilingual Travel Conversation Corpus

The translation engines used for the translation service are trained on the *Basic Travel Expressions Corpus* (ATR-BTEC) which is a collection of sentences that travel experts consider useful for people going abroad and cover a large variety of topics in travel situations like *shopping* or *stay* (Kikui et al., 2006). The multilingual corpus consists of 160K sentences for each of the 18 languages, aligned at the sentence-level. The characteristics of all ATR-BTEC corpus languages are summarized in Table 1. These languages differ largely in *word order* (SVO, SOV), *segmentation unit* (phrase, word, none), and *morphology* (poor, medium, rich). Concerning word segmentation, the corpora were preprocessed using simple tokenization tools for all European languages and language-specific word-segmentation tools for languages like Chinese, Japanese, Korean, or Thai that do not use white-space to separate word/phrase tokens. All data sets were lower-cased and punctuation marks were removed.

Table 1: Language Characteristics

Language		Order	Segments	Morphology
Arabic	(ar)	SVO	phrase	rich
Danish	(da)	SVO	words	medium
German	(de)	SVO	words	medium
English	**(en)**	**SVO**	**words**	**poor**
Spanish	(es)	SVO	words	medium
French	(fr)	SVO	words	medium
Indonesian	(id)	SVO	words	rich
Italian	(it)	SVO	words	medium
Japanese	**(ja)**	**SOV**	**none**	**poor**
Korean	(ko)	SOV	phrase	poor
Malay	(ms)	SVO	words	rich
Dutch	(nl)	SVO	words	medium
Portuguese	(pt)	SVO	words	medium
Brazilian Portuguese	(pt-b)	SVO	words	medium
Russian	(ru)	SVO	words	rich
Thai	(th)	SVO	none	none
Vietnamese	(vi)	SVO	phrase	none
Chinese	(zh)	SVO	none	none

2.2 Statistical Machine Translation Engines

Phrase-based statistical machine translation approaches continue to dominate the field of machine translation. The translation service makes use of state-of-the-art phrase-based SMT systems within the framework of feature-based exponential models containing the following features:

- Phrase translation probability
- Inverse phrase translation probability
- Lexical weighting probability
- Inverse lexical weighting probability
- Phrase penalty
- Language model probability
- Simple distance-based distortion model
- Word penalty

Table 2: Language Model Perplexity

Language	Entropy	Total Entropy	Eval Data Words	Eval Data Vocab
ar	5.73	21,663	3,780	1,067
da	5.66	17,411	3,077	884
de	5.58	16,698	2,995	910
en	**4.53**	**14,370**	**3,169**	**807**
es	5.35	15,622	2,919	943
fr	4.77	16,793	3,521	929
id	6.09	18,145	2,977	908
it	5.52	16,078	2,914	956
ja	**4.03**	**15,080**	**3,745**	**929**
ko	4.21	15,011	3,567	943
ms	6.43	19,144	2,977	909
nl	5.66	17,609	3,110	909
pt-b	5.73	16,981	2,962	932
pt	5.54	16,064	2,900	946
ru	6.20	16,040	2,587	1,143
th	5.12	20,230	3,953	738
vi	4.84	19,531	4,034	792
zh	5.11	14,748	2,887	944

Table 3: Automatic Evaluation Results

	BLEU (%)				METEOR (%)			
	en-*	*-en	ja-*	*-ja	en-*	*-en	ja-*	*-ja
ar	18.21	51.01	13.03	46.09	40.90	69.01	37.52	58.02
da	59.70	70.90	45.94	55.34	75.08	82.56	64.41	65.83
de	56.48	69.25	41.99	59.20	74.01	81.48	63.69	69.61
en	–	–	61.56	68.53	–	–	78.19	75.39
es	65.22	73.82	51.77	63.24	78.15	85.28	68.30	72.17
fr	64.69	71.04	52.36	63.16	79.28	83.05	71.14	72.82
id	48.35	59.69	40.59	57.24	66.82	75.83	62.33	69.00
it	56.80	70.43	43.45	60.77	72.41	82.96	62.35	70.70
ja	68.53	61.56	–	–	75.39	78.19	–	–
ko	37.00	58.82	69.96	85.10	57.89	75.92	83.25	89.73
ms	40.99	57.63	36.13	55.84	61.08	74.75	58.73	67.33
nl	57.46	72.85	41.43	59.70	75.88	84.52	63.42	72.19
pt-b	59.99	69.41	46.50	58.07	72.77	80.70	64.68	69.14
pt	62.81	70.25	48.24	59.20	75.65	83.32	67.38	68.32
ru	44.46	61.23	36.08	55.13	66.41	73.75	60.59	64.55
th	46.49	51.35	43.75	50.85	62.47	73.12	60.25	62.91
vi	55.18	57.42	50.86	55.07	71.04	73.98	68.67	70.81
zh	53.08	59.33	51.68	69.43	69.85	74.68	65.88	77.62

The basic framework within which all the MT systems were constructed is shown in Figure 3.

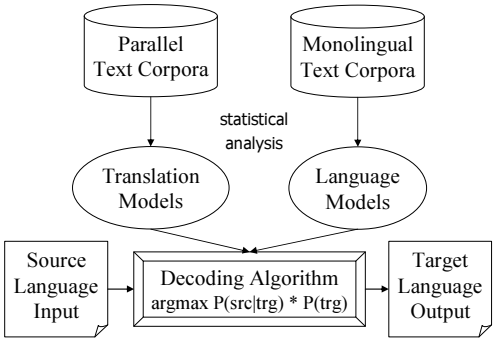

Figure 3: SMT Framework

Translation examples from the respective bilingual text corpus are aligned in order to extract phrasal equivalences and to calculate the bilingual feature probabilities. Monolingual features like the *language model probability* are trained on monolingual text corpora of the target language whereby standard word alignment and language modeling tools were used. For decoding, the *CleopATRa* decoder (Finch et al., 2007), a multi-stack phrase-based SMT decoder is used.

2.3 Evaluation

In order to get an idea of how difficult the translation tasks are, we trained standard 5gram language models on 160K sentence pairs and evaluated the *entropy* and *total entropy*, i.e., the *entropy* multiplied by *word counts*, of each language on an evaluation data set of 510 sentences each. Table 2 shows that the total entropy of European languages like Danish, German, English, Spanish, etc. does not differ much. Moreover, languages with phrasal segments and/or rich morphology like Arabic, Malay, Russian or Vietnamese have a high total entropy and thus can be expected to be more difficult to translate. This is confirmed by the translation experiments in which the evaluation data sets were translated using the servers translation engines and the translation quality was evaluated using the standard automatic evaluation metrics BLEU (Papineni et al., 2002) and METEOR (Banerjee and Lavie, 2005) where scores range between 0 (worst) and 1 (best). Besides Korean (single references only), all languages were evaluated using 16 reference translations. The evaluation results in Table 3 show that closely related language pairs like Japanese-Korean or Portuguese-Brazilian can be translated very accurately, whereas translations into languages with high total entropy are of lower quality.

3 Multilingual Speech Translation Service (MSTS)

The speech translation service[1] can be accessed via 'http://www.atr-trek.co.jp/contents_html' or using the QR code in Figure 4 that also illustrates the graphical user interface of the translation service. After connecting to the top page, the translation service is activated by selecting the "Translation" option. In order to achieve robust speech recogni-

[1]The speech translation service for Japanese⇔English on Docomo 905i mobile phones started November 2007.

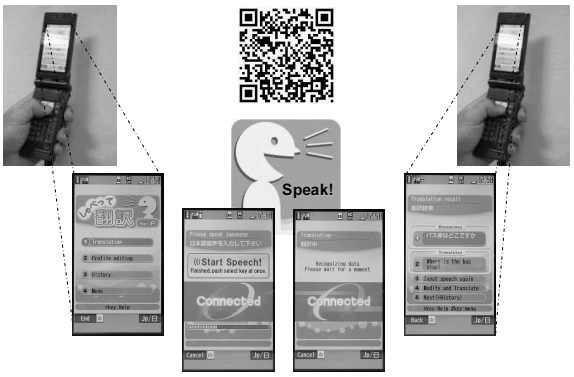

Figure 4: QR Code and GUI of MSTS

tion, the service features a push-to-talk functionality, i.e., the user (1) presses the key to start the service (2) speaks freely into the integrated microphone of the mobile phone, and (3) presses the key again after the speech input is finished. Fast and accurate front-end and back-end processing algorithms enable high-speed speech translation of the input. Both, the speech recognition results as well as the translation results are sent back to and displayed on the mobile phone.

3.1 Multilingual Speech Corpus

Similar to the statistical machine translation approach introduced in Section 2.2, the speech recognition components are based on large-sized multilingual speech corpora. For Japanese, speech recordings of 4000 speakers were collected resulting on a total of 200 hours of speech. For English, almost the same amount of speech data were collected from 500 speakers in North America (300 speakers), the UK (100 speakers), and Australia (100 speakers).

3.2 Distributed Speech Recognition

The speech interface is based on *distributed speech recognition* (DSR) that is integrated as a client-server architecture compatible with the ETSI ES 202 050 standards. The usage of *Speech Translation Markup Language* (STML) enables flexible connections between internal and external speech translation resources like speech recognition and translation servers via a network. Figure 5 illustrates the architecture of the utilized DSR system. The front-end processing includes noise suppression, feature extraction and feature parameter compression and is carried out on the mobile phone. The data stream is then sent via internet to the application service provider (ASP) for back-end processing, i.e. speech recognition and statistical machine translation. The recognition and translation results are sent back to the mobile phone for display to the user.

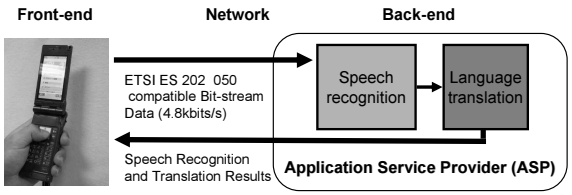

Figure 5: MSTS Architecture

4 Conclusion

This paper introduced the first commercial speech translation service in the world. State-of-the-art spoken language translation technologies (distributed speech recognition with noise suppression, multilingual statistical machine translation) are implemented into a flexible client-server architecture that covers the major languages of most countries and enables users to communicate in real environments all over the world using their own mobile phones.

5 Acknowledgments

This work is partly supported by the Grant-in-Aid for Scientific Research (C) and the Special Coordination Funds for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

Banerjee, S. and A. Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan.

Finch, A., E. Denoual, H. Okuma, M. Paul, H. Yamamoto, K. Yasuda, R. Zhang, and E. Sumita. 2007. The NICT/ATR Speech Translation System for IWSLT 2007. In *Proc. of the IWSLT*, pages 103–110, Trento, Italy.

Kikui, G., S. Yamamoto, T. Takezawa, and E. Sumita. 2006. Comparative study on corpora for speech translation. *IEEE Transactions on Audio, Speech and Language Processing*, 14(5):1674–1682.

Papineni, K., S. Roukos, T. Ward, and W. Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proc. of the 40th ACL*, pages 311–318, Philadelphia, USA.