POSSLT: A Korean to English Spoken Language Translation System

Donghyeon Lee, Jonghoon Lee, Gary Geunbae Lee

Department of Computer Science and Engineering

Pohang University of Science & Technology (POSTECH)

San 31, Hyoja-Dong, Pohang, 790-784, Republic of Korea

{semko, jh21983, gblee}@postech.ac.kr

Abstract

The POSSLT [1] is a Korean to English spoken language translation (SLT) system. Like most other SLT systems, automatic speech recognition (ASR), machine translation (MT), and text-to-speech (TTS) are coupled in a cascading manner in our POSSLT. However, several novel techniques are applied to improve overall translation quality and speed. Models used in POSSLT are trained on a travel domain conversational corpus.

1 Introduction

Spoken language translation (SLT) has become more important due to globalization. SLT systems consist of three major components: automatic speech recognition (ASR), statistical machine translation (SMT), text-to-speech (TTS). Currently, most of SLT systems are developed in a cascading method. Simple SLT systems translate a single best recognizer output, but, translation quality can be improved using the N-best hypotheses or lattice provided by the ASR (Zhang et. al., 2004; Saleem et. al., 2004).

In POSSLT, we used an N-best hypothesis re-ranking based on both ASR and SMT features, and divided the language model of the ASR according to the specific domain situation. To improve the Korean-English SMT quality, several new tech-

[1] POSSLT stands for POSTECH Spoken Language Translation system

niques can be applied (Lee et. al., 2006-b). The POSSLT applies most of these techniques using a preprocessor.

2 System Description

The POSSLT was developed by integrating ASR, SMT, and TTS. The system has a pipelined architecture as shown in Fig. 1. LM loader, preprocessor and re-ranking module are newly developed to improve the translation quality and speed for POSSLT.
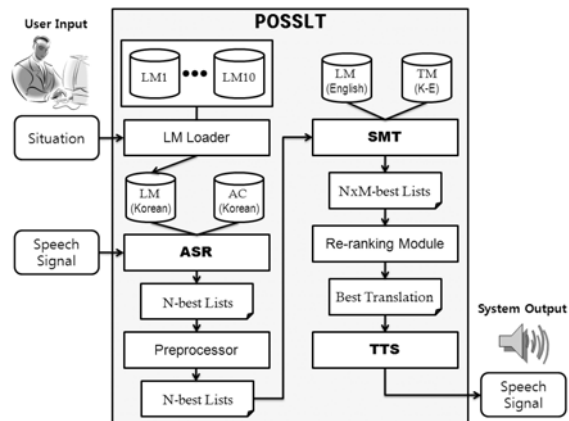

Figure 1: Overview of POSSLT

2.1 ASR

The system used HTK-based continuous speech recognition engine properly trained for Korean. The acoustic model, lexical model and language model of Korean are trained for conversational corpus. The phonetic set for Korean has 48 phoneme-like-units, and we used three-state tri-phone hidden Markov models and trigram language mod-

els. Pronunciation lexicons are automatically built by a Korean grapheme-to-phoneme (G2P) tool (Lee et. al., 2006-a). We used an *eojeol*[2] as a basic recognition unit for lexical and language models, because an *eojeol*-based recognition unit has the higher accuracy than the morpheme-based one. The ASR produces the N-best hypotheses determined through the decoding process, which are used as the input of SMT.

2.2 SMT

We implemented a Korean-English phrase-based SMT decoder based on Pharaoh (Koehn, 2004). The decoder needs a phrase translation model for the Korean-English pair and a language model for English. We used the Pharaoh training module and GIZA++ (Och and Ney, 2000) to construct the phrase translation table. For language modeling, SRILM toolkit (Stolcke, 2002) was used to build a trigram language model.

2.3 TTS

We used Microsoft SAPI 5.1 TTS engine for English TTS. The final best translation is pronounced using the engine.

2.4 LM Loader

In cascading SLT systems, SMT coverage depends on the used ASR. In order to increase the ASR coverage, our system loads and unloads the ASR language models dynamically. In our system which uses a travel corpus, language models are built for ten domain situation categories such as an airport, a hotel, a shopping, etc. Besides user utterances, user selection of the situation is needed as an input to decide which language model have to be loaded in advance. By using the divided language models, many benefits such as fast decoding, higher accuracy and more coverage can be obtained.

2.5 Preprocessor

In the Korean-English SMT task, there have been developed several techniques for improving the translation quality such as changing spacing units into morphemes, adding POS tag information, and deleting useless words (Lee et. al., 2006-b).

However, for these techniques, Part-Of-Speech (POS) tagger is needed. If the final analyzed form of an *eojeol* (in the form of a sequence of morphemes plus POS tags) is defined as a word in the ASR lexicon, the transformed sentences are directly generated by the ASR only, so POS tagger errors can be removed from the system. Preprocessor also removes useless words in SMT in the transformed sentences produced by the ASR.

2.6 Re-ranking Module

We implemented a re-ranking module to make a robust SLT system against the speech recognition errors. The re-ranking module uses several features: ASR acoustic model scores, ASR language model scores, and SMT translation scores. Finally, the re-ranking module sorts the N-best lists by comparing the total scores.

Acknowledgements

References

A. Stolcke. 2002. *SRILM – An Extensible Language Modeling Toolkit*. Proc. of ICSLP.

F. J. Och and H. Ney. 2000. *Improved statistical alignment models*. Proc. of 38th Annual Meeting of the ACL, page 440-447, Hongkong, China, October 2000.

Jinsik Lee, Seungwon Kim, Gary Geunbae Lee. 2006-a. *Grapheme-to-Phoneme Conversion Using Automatically Extracted Associative Rules for Korean TTS System*. Proc. of Interspeech-ICSLP.

Jonghoon Lee, Donghyeon Lee, Gary Geunbae Lee. 2006-b. *Improving Phrase-based Korean-English Statistical Machine Translation*. Proc. of Interspeech-ICSLP.

P. Koehn. 2004. *Pharaoh: A Beam Search Decoder for Phrase-based Statistical Machine Translation Models*. Proc. of AMTA, Washington DC.

R. Zhang, G. Kikui, H. Yamamoto, T. Watanabe, F. Soong, and W. K. Lo. 2004. *A unified approach in speech-to-speech translation: Integrating features of speech recognition and machine translation*. Proc. of Coling 2004, Geveva.

S. Saleem, S. Chen Jou, S. Vogel, and T.Schultz. 2004. *Using word lattice information for a tighter coupling in speech translation systems*. Proc. of ICSLP 2004, Jeju, Korea.

---

[2] *Eojeol* is a spacing unit in Korean and typically consists of more than one morpheme.