Hierarchical Phrase-Based MT for Phonetic Representation-Based Speech Translation

Zeeshan Ahmed†, Jie Jiang‡, Julie Carson-Berndsen†, Peter Cahill† and Andy Way‡

†CNGL, School of Computer Science and Informatics, University College Dublin, Ireland
zeeshan.ahmed@ucdconnect.ie,{julie.berndsen,peter.cahill}@ucd.ie

‡Applied Language Solutions, Delph, Greater Manchester, United Kingdom
{jie.jiang, andy.way}@appliedlanguage.com

Abstract

The paper presents a novel technique for speech translation using hierarchical phrased-based statistical machine translation (HPB-SMT). The system is based on translation of speech from phone sequences as opposed to conventional approach of speech translation from word sequences. The technique facilitates speech translation by allowing a machine translation (MT) system to access to phonetic information. This enables the MT system to act as both a word recognition and a translation component. This results in better performance than conventional speech translation approaches by recovering from recognition error with help of a source language model, translation model and target language model. For this purpose, the MT translation models are adopted to work on source language phones using a grapheme-to-phoneme component. The source-side phonetic confusions are handled using a confusion network. The result on IWLST'10 English-Chinese translation task shows a significant improvement in translation quality. In this paper, results for HPB-SMT are compared with previously published results of phrase-based statistical machine translation (PB-SMT) system (Baseline). The HPB-SMT system outperforms PB-SMT in this regard.

1 Introduction

Hierarchical phrase-based statistical machine translation (HPB-SMT) is a statistical machine translation (SMT) approach which uses syntactic information of a language-pair for translation. The translation hypotheses are generated based on hierarchical rules in CYK parsing (Chiang, 2007). Although, the HPB-SMT makes use of syntactic information but it does not require syntactically annotated resources during training as syntactic information is automatically inferred from the training data. This makes the training process fully automatic without any knowledge of language-pair.

MT systems utilizing syntactic knowledge have had significant success in recent year for text based translation tasks. The strength of a system using syntactic knowledge (Chiang, 2007; Zollmann and Venugopal, 2006; Melamed, 2004; Yamada and Knight, 2002; Weese et al., 2011) over the simpler phrase-based SMT (PB-SMT) (Koehn et al., 2003; Och and Ney, 2004) is their power of translation between dissimilar language pair e.g. English-Chinese. They are also being applied for the speech translation task as highlighted by the significant number of systems participating in recent IWSLT workshops (Paul et al., 2010; Federico et al., 2011) which are in one-way or another making use of syntactic information during translation. Much of the improvement in these systems is mostly because of an improvement in the text based translation technique while keeping the automatic speech recognizer (ASR) as a black-box; providing 1-best word output, confusion networks or lattices as input to the MT system. The technique proposed in this paper goes one step further. It uses the phonetic knowledge from ASR to help improve the speech translation quality for HPB-SMT.

The most commonly followed method for developing a speech translation system is the cascade approach. In this approach, ASR, MT and speech

synthesis systems are used as black-boxes for each other. The basic unit of information sharing between these components is word i.e the speech is fed into ASR to obtain 1-best,n-best (Zhang et al., 2004) lists, word lattices (Matusov and Ney, 2011; Matusov et al., 2005) or confusion networks (Bertoldi et al., 2008b) then the recognized output is translated into target language using the MT component. The target speech is then synthesized using a speech synthesis component. Except for input-output, there is no information sharing between these components. This approach is straight forward to implement and improvement can be obtained by individually improving each component. But, there are still some drawbacks for cascade approach.

- Most of the linguistic information (Phrases, Syntax etc.) about the language is used on the MT side. It becomes very late for an MT component to apply such linguistic knowledge on word-level output. Which, if applied at the acoustic-phonetic level could cause a significant improvement in system accuracy.

- Recognition errors caused by the ASR are propagated through the MT component to the target speech. It is difficult to recover from error at word-level even when word lattices or confusion networks are employed.

- Tuning of the three components together is difficult. As each of the components is trained on a different corpus with the assumption that domain is coherent for each component.

Another completely different approach to the cascade model is the tightly integrated model. In the tightly integrated model, the ASR and MT components are tightly integrated such that speech recognition and translation are done in single step. A finite state transducer (Bangalore and Riccardi, 2000; Casacuberta et al., 2001; Mathias and Byrne, 2006) is widely used for this task. This translation approach is similar to speech recognition except that the system outputs text in the target language. This approach has the same problem as speech recognition for large vocabularies.

All of the previous studies in speech translation use the word as a basic unit for translation. Recently, an approach for speech translation from phonetic-representation was proposed in (Jiang et al., 2011). In this approach, PB-SMT is used for translation of the source language from a phone sequence directly into the target text. The approach uses a confusion network (CN) to deal with phonetic confusions. It has outperformed the MT system operating on word input as highlighted by the results presented in the paper.

1.1 Motivation for HPB-SMT

In this paper, a new paradigm for phonetic representation-based speech translation is presented which uses a HPB-SMT technique. For the systems working on text-based translation, HPB-SMT has been shown to perform better than PB-SMT on dissimilar language pairs (Chiang, 2007).

The main strength of HPB-SMT systems over PB-SMT lies in their ability to perform better lexicalized reordering and translation of discontinuous phrases (Lopez, 2008). The performance comparison between PB-SMT, HPB-SMT and Syntax Augmented Machine Translation (SAMT) has been presented in (Zollmann et al., 2008). The performance is analysed on Chinese–English and Arabic–English translation under different language model size conditions. It was shown that the HPB-SMT consistently performs better than PB-SMT for Chinese–English translation even when a larger language model is used which actually favors PB-SMT reordering model. However, this improvement is not consistent for Arabic-English translation which results in conclusion that the HPB-SMT performs better than PB-SMT for language pairs which are non monotonic i.e. long reordering ranges are required for those languages which is the case with Chinese-English translation. Furthermore, (Zollmann et al., 2008) found that PB-SMT could not produce 22% of the translation generated by HPB-SMT for the Chinese-to-English NIST MT06 test set using forced translation, which highlights HPB-SMT's ability to translate discontinuous phrases.

However, taking all of the advantages of text-based translation, the motivation here is much more centric towards the speech translation task. The HPB-SMT uses parsing based computational models for translation which facilitates the application of syntax-based source language model over an in-

put without any additional cost. It has been shown in (Ahmed et al., 2012) that syntactic parsing as a language model over phonetic space can result in improvement in word recognition accuracy as well as syntactic accuracy that is likely to favor speech translation task. The source side language features have also shown to be very helpful in text-based HBP-SMT e.g. in (Du and Way, 2010), the role of source side reordering of DE grammatical structure in Chinese is analyzed for better translation quality between Chinese-English. Therefore, the HPB-SMT offers number of advantages for translation of speech from phone sequences over PB-SMT.

- The phone-based translation using PB-SMT performs well when there are limited confusions in the input confusion network. The translation quality degrades with the increase of confusions in the confusion network. The main reason for this is the absence of source language model constraints. HPB-SMT performs better than PB-SMT in this regard because it has the inherent power of applying a syntactic language model in the form of hierarchical phrase rules. The parsing of source language together with n-gram target language modelling effectively controls the search space of the decoder in the case of dense confusion networks.
- It allows the application of hierarchical syntactic knowledge at the phonetic level which is impractical for PB-SMT decoder or FST based approaches in general.

The rest of the paper is organized as follows. The next section presents the detailed description of phone translation system based on PB-SMT as outlined in (Jiang et al., 2011). The proposed phone translation system based on HPB-SMT is presented in section 3. The evaluation of systems is presented in 4 and conclusions are finally drawn in section 5.

2 Phrase-Based Phone MT

This section describes in detail the phone MT approach originally presented in (Jiang et al., 2011). The overall system architecture is presented in figure 1. In this architecture, the ASR and MT system are neither tightly integrated as in the case of FST based approach nor loosely coupled as in the case of cascade approach. Hence, it can be referred to as a semi-integrated approach for speech translation. In this approach;

- the role of ASR is reduced such that the ASR task is just to recognized phones of the language.
- the role of MT is increased such that all the major linguistic analyses (phonetic modelling, language modelling ,translation etc.) are performed during the translation process.

The approach has multiple advantages over the conventional loosely and tightly integrated approaches.

- MT uses phonetic representation of speech for translation. This helps in homograph disambiguation (e.g. "read" has the same orthographic representation for present and past tense but is pronounced differently).
- Out-of-vocabulary (OOV) words are critical to handle in the cascade approach as it requires ASR and MT to use their own approaches for OOV handling. The approach offers the single point for handling OOV words. The MT decoder can also handle ASR OOV (Ahmed and Carson-Berndsen, 2010) words in addition to handling MT OOV words and recognition errors.
- the ASR can be general purpose while the domain tuning can be performed at MT level.
- Furthermore, speech translation is not just a translation of language but also a translation of different phenomena present in speech e.g. voice characteristics, speaker expressions, tone or prosody etc. In our hypothesis, phonetic representation-based speech translation is an ideal approach for translation of these phenomena as phones are the basic unit for managing these characteristics that can carry such information to target language effectively.

2.1 How It Works

To perform the translation from phone sequence to target language, the MT system must be aware of

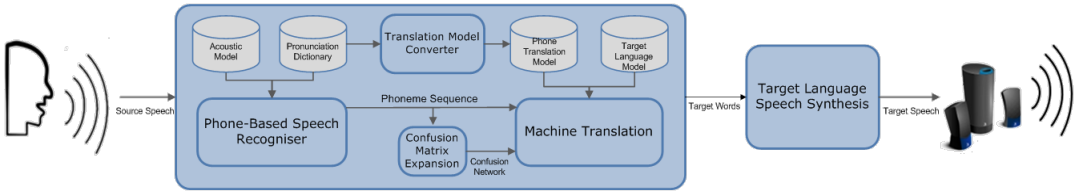

Figure 1: Phone-to-word SMT for speech translation using phonetic representation

Source Entry	Target Entry
suggest you go	最 好 去
suggest you have	建议 你 要
suggest you look	建议 你
suggest you take	建议 你 坐
suggest you visit	还 建议 你 参观
suggest you	你 最 好

(a) Original Translation Model

Source Entry	Target Entry
S AH G JH EH S T Y UW G OW	最 好 去
S AH G JH EH S T Y UW HH AE V	建议 你 要
S AH G JH EH S T Y UW L UH K	建议 你
S AH G JH EH S T Y UW T EY K	建议 你 坐
S AH G JH EH S T Y UW V IH Z IH T	还 建议 你 参观
S AH G JH EH S T Y UW	你 最 好

(b) Adapted Translation Model

Table 1: Original and adapted translation model for PB-SMT using G2P.

the source language phonetic knowledge. For this purpose, the MT translation models are adapted to include phonetic knowledge such that the MT component can handle input in the form of a phone sequence rather than words. The ASR system also needs to be adapted so that it should output a phone sequence rather than words.

2.2 Obtaining Phonetic Knowledge

The phone sequence of speech can be derived either by using general purpose phone recognizer or by converting the word recognition output into phones. The difference here is the language model applied in the recognition process which actually affects the phone recognition rate for the subsequent MT process. For the first approach, a higher order phone language model is preferred for a better phone recognition rate (Bertoldi et al., 2008a). This approach can be beneficial for the languages which do not have diversity in pronunciation system and do not differ considerably in orthographic and pronunciation systems. While, in the second approach, the phonetic knowledge can be used as an aid to word-based translation. The languages like English, which have vast diversity in pronunciation systems can benefit from this approach. In the phone MT work presented in (Jiang et al., 2011) and the HPB-SMT phone MT work presented in this paper, the second approach is only examined in order to compare with systems that operate on the word-level, but the method can be generally applied to any speech recognizer that outputs phone sequences.

2.3 Adapting Translation Models

The training process for generating translation models for phone MT is same as that of the word MT i.e. the system is trained at word-level on parallel corpus. After the normal training is completed, the translation table and re-ordering table are adapted to include source language phonetic knowledge. The source side entries of both tables are converted from words to phones using a grapheme-to-phoneme (G2P) converter. The table 1.a and 1.b show the snippet of original translation table for word MT and adapted translation table for phone MT respectively.

2.4 The Use of the G2P Converter

The G2P converter plays a vital role in phone-based MT system. It is used for both adaptation of translation model as well as conversion of recognized output into phone sequence. Simple approach to G2P converter is to use a pronunciation dictionary but dictionary does not cover all of the words of a language specially proper nouns. Therefore, a phrase-based log-linear translation model is used for G2P conversion as described in (Jiang et al., 2011).

2.5 Dealing with Phonetic Confusions

The phone MT system under-performs while operating at the phone-level as compared to word MT. It is because the phone-level input broadens the search space of MT decoder causing it to make wrong decisions. However, the merit of the phone MT is the flexibility to incorporate phonetic information. For example, considering the error-prone nature of ASR, multiple phone choices can be provided to phone MT if the information is available about which phones are closer to others based on the recognizer outputs. This information is usually represented by phone confusion matrix (PCM) and easily encoded by phones in the form of a phone confusion network (PCN). PCN can be well-handled by state-of-the-art MT engines. The same approach of PCN generation is followed here as presented in (Jiang et al., 2011). The same terminology of "Confusion Matrix Enhanced Phone Translation" (CMEPT) is also used for the system using PCM information.

3 Hierarchical Phrase-Based Phone MT

The HPB-SMT translation model is based on weighted synchronous context free grammar (SCFG) (Aho and Ullman, 1969). The SCFG is similar to CFG except that the rewrite rules contain two right-hand sides corresponding to source and target language with aligned non-terminal. The formal structure of rewrite rules is shown below

$$X \rightarrow \langle \gamma, \alpha, \sim \rangle \;\; \gamma, \alpha \in \{N \cup W\}$$

where, γ and α are the strings of terminal and non-terminal symbols, N is a set of non-terminal, X is a non-terminal, W is a set of words or terminals and $\sim$ is a one-to-one alignment between the non-terminal in γ and corresponding non-terminal in α.

To adapt the model to work from phonetic representation, the source side entry of the right-hand side in the hierarchical rule table is transformed into a phonetic representation using a rule table phonetic transformation function T which is defined as; $\bigvee w \in \gamma$

$$T(w) = \begin{cases} Phonetic(w) & \text{if } w \in W \\ w & \text{otherwise} \end{cases}$$

$Phonetic : W \rightarrow P$ is a function defined over set of words W and their pronunciations P. This function is modelled with G2P converter.

Normally, a word can have multiple pronunciations in that case the $Phonetic(w)$ does not represent a function. Such a case is avoided here. It is because it is impractical to have a phrase rule with all possible pronunciation variations. For example, if a phrase contains n number of words and each word in that phrase have m number of pronunciation then the transformed rule can have m^n corresponding rules. Therefore, only base form of a word pronunciation provided by G2P is used for rule transformation. The case of multiple pronunciations is easily handled by phone confusion network at run-time.

An example of an original and a transformed rule table is shown in table 2. Similar kind of transformation can be performed on target-side entry of rule with target language G2P, if it is intended to get the phonetic form of target words for speech synthesis purpose.

3.1 Confusion Network Translation

All of the state-of-the-art syntax-based decoders for machine translation (Dyer et al., 2010; Weese et al., 2011) have a facility to operate on word lattices (Dyer et al., 2008). A confusion network is also a type of lattice that has the peculiarity that each path from the start node to the end node goes through all the other nodes and may contain an additional arc labelled $*delete*$ to skip unwanted item in an input string. To translate confusion networks, two rules are further introduced in the hierarchical phrase grammar as shown below.

$$X \rightarrow \langle X \;\; *delete*, X \rangle$$
$$X \rightarrow \langle *delete* \; X, X \rangle$$

4 Experiment and Evaluation

The HPB-SMT phone MT system is evaluated on the IWSLT 2010 English-Chinese corpus[1]. The corpus contains spoken dialogues related to the travel domain. The selected training set contains 71,725 parallel sentences pairs which is used to train both a translation model and a language model. The development set contains 498 sentences that is used

[1] http://iwslt2010.fbk.eu/

Source Entry	Target Entry
your key [X,1] is	您 的 钥匙 [X,1] 是
your key [X,1] luggage	您 房间 的 钥匙 [X,1] 行李
your key [X,1] porter	您 的 钥匙 [X,1] 行李 搬运工
your key [X,1] room	您 的 钥匙 [X,1] 房间
your key [X,1] someone	您 房间 的 钥匙 [X,1] 人
your key [X,1] will [X,2]	您 的 钥匙 [X,1] 会 [X,2]

(a) Original Translation Model

Source Entry	Target Entry
Y UH R K IY [X,1] IH Z	您 的 钥匙 [X,1] 是
Y UH R K IY [X,1] L AH G IH JH	您 房间 的 钥匙 ...
Y UH R K IY [X,1] P AO R T ER	您 的 钥匙 [X,1] ...
Y UH R K IY [X,1] R UW M	您 的 钥匙 [X,1] ...
Y UH R K IY [X,1] S AH M W AH N	您 房间 的 钥匙 ...
Y UH R K IY [X,1] W AH L [X,2]	您 的 钥匙 [X,1] ...

(b) Adapted Translation Model

Table 2: Original and adapted translation model for Hierarchical MT using G2P.

for tuning MT parameters, while PCM is extracted from another development set of 2,060 1-best outputs and transcriptions selected from IWSLT English–Chinese DIALOG development sets. The test set comes from 1-best ASR outputs with a WER of 17.9% and contains 251 sentences. It remains entirely unseen during the training process. The development and test sets contain 7 references. The CMU[2] dictionary is used for training G2P module which is then used for deriving phonetic form of the words.

The main objective of the evaluation is to compare the performance of phonetic representation-based speech translation between PB-SMT (baseline) and HPB-SMT (proposed). Three different systems are developed for each paradigm (PB-SMT and HPB-SMT) for comparison purposes; word MT, phone MT and CMEPT. The baseline PB-SMT is built using the Moses (Koehn et al., 2007) toolkit. Moses is also used for training HPB-SMT model. The JOSHUA (Weese et al., 2011) decoder is then used for decoding because of the facility it provides for lattice translation.

The CMEPT systems for both paradigms are controlled by a confusion threshold (CT) parameter i.e all of the phone confusions that have a score less than the CT value are pruned during translation. The results for CMEPT systems are presented for different CT values to highlight the gradual improvement of the CMEPT systems.

All of the word-level and phone-level PB-SMT settings are similar to the one described in (Jiang et al., 2011). For HPB-SMT systems, the following training steps are followed.

- standard HPB-SMT training is performed on the parallel corpus at word-level.
- the 5-gram language model is trained on the target language using SRILM (Stolcke, 2002).
- the features used for word-level systems are phrase probability $P(e|f)$, lexical probability $lexprob(e|f)$, inverse lexical probability $lexprob(f|e)$, word penalty and language model.
- the word-level system feature weights are optimized on the development set using MERT (Och, 2003) with BLEU (Papineni et al., 2002) as evaluation criteria.
- After optimising the feature weights, rule table phonetic transformation is performed to make the model work on phonetic input.
- Now the model works on phone-level input. The feature weights are further optimized for phone MT using one more MERT operation on the development set using a span-limit[3] reasonable for phone sequence input. For this purpose, the development set is also transformed into the phonetic form using G2P. To work on the confusion network, each input in the development set needs to be transformed into a phone confusion network as described in section 2.5 prior to MERT so that the PCN transition parameter (confusion network confidence scores) can also be optimized.

The table 3 shows the BLEU scores for all of the systems. The BLEU on correct text is also presented for reference. It should be noted that overall performance of the HPB-SMT system is better than PB-SMT as has been previously mentioned in

[2]http://www.speech.cs.cmu.edu/cgi-bin/cmudict

[3]In syntax-based machine translation, span-limit refers to the maximum number of words a rule can cover.

System Type	CT	PB-SMT (Baseline)	HPB-SMT (Proposed)
Correct text	-	34.86	37.48
Word MT	-	29.60	30.57
Phone MT	-	28.43	29.90
CMEPT	0.01	30.14	31.81
	0.008	**30.87**	31.84
	0.006	30.78	31.97
	0.004	30.43	31.94
	0.003	30.43	**32.47**

Table 3: BLEU score for IWSLT 2010 Spoken dialogue Translation task for PB-SMT and HPB-SMT system.

literature (Chiang, 2007). The phone-level MT system under-performs for both categories (PB-SMT, HPB-SMT) with respect to their word-level systems. The best performance for PB-SMT achieved is for its CMEPT system giving 30.87 BLEU at CT value of 0.008. It outperforms word-based PB-SMT by 1.27% absolute (4.29% relatively) BLEU points and phone-based PB-SMT by 2.44% absolute (8.58% relative) BLEU points.

On the other-hand, the best performance achieved for HPB-SMT is also for its CMEPT system giving 32.47 BLEU at CT value of 0.003. It outperforms word-based HPB-SMT system by 1.9% absolute (6.21% relative) BLEU points and word-based PB-SMT by 2.87% absolute (9.38% relative) BLEU points. The BLEU score of 32.47 is the best result obtained during the experiment.

4.1 Discussion

The main reason for the better performance of the CMEPT system is definitely recognition error recovery at MT level using confusion network. But, CMEPT system with hierarchical phrase-based modelling performance is even better than CMEPT with simple phrase-based modelling. One of the reasons for improved performance is the confusion network as for the case of PB-SMT, and the other is hierarchical syntax translation rules. In the following sections, the role of each of the technique is discussed in detail.

4.1.1 The Role of Confusion Network

The results showed that phone MT did not performed as well as their corresponding word MT systems for both paradigms. The major reason for this is the broader search space presented by phonetic forms of the words to the decoder. However, the use of confusion network to deal with phonetic confusion has played an important role in both PB-SMT and HPB-SMT systems.

The following example illustrates why the CMEPT system performance is better than 1-best word and phone outputs.

Correct: under one thousand yen
ASR: and er one thousand yen

HPB Word MT:
Translation: 那 一千 日元
literal: That one thousand yen

HPB Phone MT(CT=0.003):
Translation: 一千 日元 以下
literal: One below thousand yen

During the recognition process the word "under" was mis-recognized as "and er". The mis-recognition causes the sentence to be translated incorrectly for the word "under". While, looking closely at the phonetic form of both of words; "under" and "and er", which are /AH N D ER/ and /AE N D ER/ respectively, it reveals that the phonetic forms are almost the same except for the starting phones. The CMEPT recovers from this error because of the information provided by PCM about the phone /AE/ and the phones which are acoustically and phonetically similar to /AE/. The translation provided by CMEPT system is literally better than one provided by word MT. The phone confusion network for this example is shown in figure 2 with selected path highlighted in red.

4.1.2 The Role of Syntax

The syntactic information also played an important role in better translation quality for HPB-SMT system overall. This fact is evident by the results obtained for HPB-SMT systems. For the CMEPT-HPB system, the syntactic information proved to be very beneficial as it impose tight constraint over con-

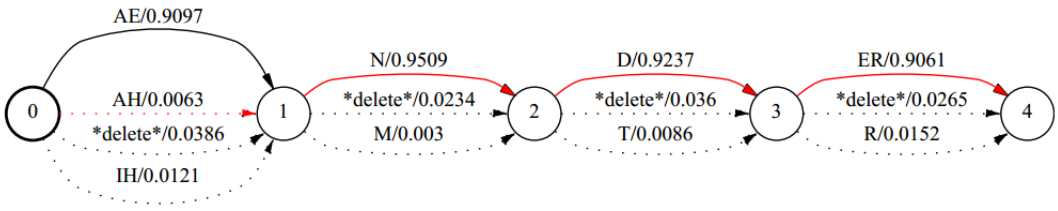

Figure 2: Example Phone Confusion Network

fusion network parsing to avoid mis-recognition at lower confusion thresholds. The PB-SMT worked well when confusions were within limits. Its accuracy started degrading at CT values lower than 0.008. It is mainly because of the missing source language model constraints. This problem was overcome by the HPB-SMT system using hierarchical phrase rules which act as a source language model during translation.

The following example dialogue from test set illustrates the role of syntax during translation.

Correct: do you have any bean cake?
ASR: do you have any been cakes

PB Phone MT (CT = 0.008):
Translation: 有 蛋糕
literal: Has the cake

HPB Word MT:
Translation: 有 被 烧饼
literal: Has by the bean cake

HPB Phone MT (CT = 0.003):
Translation: 你 有 烧饼
literal: You have the bean cake

The ASR makes an error in the ending phrase recognising "bean cake" as "been cakes". It is to be noted that similar to confusion network example shown previously, the pronunciation of "bean" and "been" is exactly the same i.e. /B IY N/. Just because of this fact, the PB-SMT system is not able to give better translation. Even though, the example translation presented above is for the best CMEPT-PB-SMT system at CT value of 0.008. On the other-hand, CMEPT-HPB system handles this with hierarchical syntax rules which provide the translation that is literally very close to original sentence.

5 Conclusion

The paper presented a new paradigm for phonetic representation-based speech translation using hierarchical phrase-based machine translation technique. The phonetic representation-based speech translation also called semi-integrated approach to speech translation is a technique of speech translation where translation is performed from phone sequence of speech rather than word sequence. In this way, the machine translation system also act as a word recognition system in addition to translation system.

This paper highlighted the role of syntax in phonetic representation-based speech translation. It was presented that syntactic parsing of source language and syntactic constraints of hierarchical phrase rules over confusion network resulted in better translation quality over a previously published results of a system which used a PB-SMT (Jiang et al., 2011). The results presented in the paper showed that HPB-SMT has a improvement of 9.38% (relative) BLEU points than baseline PB-SMT. The main source of improvement in translation quality is error recovery in ASR recognition output.

The role of source language model (syntactic or n-gram) is very important in phonetic representation-based speech translation. The missing source language model is the main reason for low performance of PB-SMT system. In future, the plan is to use additional source side n-gram language model feature for further improvement in source side recognition. Furthermore, it is also desirable to investigate the role of full syntactic parsing against the parsing offered by HPB.

Acknowledgments

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) at University College Dublin. The opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Science Foundation Ireland.

References

Zeeshan Ahmed and Julie Carson-Berndsen. 2010. Modeling Pronunciation of OOV Words for Speech Recognition. In *Thirteenth Australasian International Conference on Speech Science and Technology*, Melbourne, Australia.

Zeeshan Ahmed, Peter Cahill, and Julie Carson-Berndsen. 2012. Phonetically aided syntactic parsing of spoken language. In *Proceedings of the 11th Conference on Natural Language Processing (KONVENS 2012)*, Vienna, Austria, September.

A. V. Aho and J. D. Ullman. 1969. Syntax directed translations and the pushdown assembler. *Journal of Computer and System Sciences*, 3(1):37--56, February.

Srinivas Bangalore and Giuseppe Riccardi. 2000. Stochastic finite-state models for spoken language machine translation. In *ANLP-NAACL 2000 Workshop: Embedded Machine Translation Systems*, EmbedMT '00, pages 52--59, Stroudsburg, PA, USA. Association for Computational Linguistics.

Nicola Bertoldi, Marcello Federico, Giuseppe Falavigna, and Matteo Gerosa. 2008a. Fast speech decoding through phone confusion networks. In *INTERSPEECH*, pages 2094--2097, Brisban, Australia.

Nicola Bertoldi, Richard Zens, and Marcello Federico. 2008b. Efficient speech translation through confusion network decoding. In *IEEE Transactions on Audio, Speech, and Language Processing*, pages 1696 -- 1705, Honolulu, HI.

F Casacuberta, D Llorens, C Martinez, S Molau, F Nevado, H Ney, M Pastor, D Pico, A Sanchis, E Vidal, and JM Vilar. 2001. Speech-to-speech translation based on finite-state transducers. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'01.*, pages 613--616. IEEE Press.

David Chiang. 2007. Hierarchical phrase-based translation. *Computational Linguitics*, 33:201--228.

Jinhua Du and Andy Way. 2010. The impact of source-side reordering on hierarchical phrase-based smt. In *Proceedings of the 14th Annual Conference of the European Association for Machine Translation (EAMT 2010)*, pages 82--89, Saint-Raphaël, France.

Christopher Dyer, Smaranda Muresan, and Philip Resnik. 2008. Generalizing word lattice translation. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, Columbus, Ohio, USA.

Chris Dyer, Adam Lopez, Juri Ganitkevitch, Johnathan Weese, Ferhan Ture, Phil Blunsom, Hendra Setiawan, Vladimir Eidelman, and Philip Resnik. 2010. cdec: A decoder, alignment, and learning framework for finite-state and context-free translation models. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Marcello Federico, Luisa Bentivogli, Michael Paul, and Sebastian Stuker. 2011. Overviewof the iwslt 2011 evaluation campaign. In *International Workshop on Spoken Language Translation (IWSLT'11)*, San Francisco, USA.

Jie Jiang, Zeeshan Ahmed, Julie Carson-Berndsen, Peter Cahill, and Andy Way. 2011. Phonetic representation-based speech translation. In *13th Machine Translation Summit*, Xiamen , China.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, NAACL '03, pages 48--54, Stroudsburg, PA, USA. Association for Computational Linguistics.

Philipp Koehn, Alexandra Birch Hieu Hoang, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Annual Meeting of the Association for Computational Linguistics(ACL), demonstration session*, Prague, Czech Republic, June.

Adam Lopez. 2008. Tera-scale translation models via pattern matching. In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*, COLING '08, pages 505--512, Stroudsburg, PA, USA. Association for Computational Linguistics.

Lambert Mathias and William Byrne. 2006. Statistical phrase-based speech translation. In *IEEE International Conference Acoustics, Speech and Signal Processing (ICASSP)*.

Evgeny Matusov and Hermann Ney. 2011. Lattice-based asr-mt interface for speech translation. *IEEE Transactions on Audio, Speech, and Language Processing*, May.

Evgeny Matusov, Hermann Ney, and Ralph Schluter. 2005. Phrase-based translation of speech recognizer

word lattices using log-linear model combination. In *IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 110 -- 115.

I. Dan Melamed. 2004. Statistical machine translation by parsing. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 653--660, Barcelona, Spain, July.

Franz Josef Och and Hermann Ney. 2004. The alignment template approach to statistical machine translation. *Computational Linguistics*, 30(4):417--449, December.

Franz Josef Och. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, pages 160--167, Stroudsburg, PA, USA. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *40th Annual meeting of the Association for Computational Linguistics*, pages 311--318, Philadelphia, PA, USA.

Michael Paul, Marcello Federico, and Sebastian Stuker. 2010. Overview of the iwslt 2010 evaluation campaign. In *International Workshop on Spoken Language Translation (IWSLT'10)*, Paris, France.

Andreas Stolcke. 2002. SRILM - An Extensible Language Modeling Toolkit. In *International Conference on Spoken Language Processing*, Denver, Colorado.

Jonathan Weese, Juri Ganitkevitch, Chris Callison-Burch, Matt Post, and Adam Lopez. 2011. Joshua 3.0: Syntax-based machine translation with the thrax grammar extractor. In *6th Workshop on Statistical Machine Translation*, pages 478--484, Edinburgh, Scotland.

Kenji Yamada and Kevin Knight. 2002. A decoder for syntax-based statistical mt. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 303--310, Stroudsburg, USA.

Ruiqiang Zhang, Genichiro Kikui, Hirofumi Yamamoto, Taro Watanabe, Frank Soong, and Wai Kit Lo. 2004. A unified approach in speech-to-speech translation: integrating features of speech recognition and machine translation. In *Proceedings of the 20th international conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Andreas Zollmann and Ashish Venugopal. 2006. Syntax augmented machine translation via chart parsing. In *Proceedings of the Workshop on Statistical Machine Translation*, pages 138--141, Stroudsburg, PA, USA.

Andreas Zollmann, Ashish Venugopal, Franz Och, and Jay Ponte. 2008. A systematic comparison of phrase-based, hierarchical and syntax-augmented statistical mt. In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*, COLING '08, pages 1145--1152, Stroudsburg, PA, USA. Association for Computational Linguistics.