ISCA Archive
http://www.isca-speech.org/archive

Spoken Language Technologies for
Under-Resourced Languages
(SLTU-2010), Penang, Malaysia

EXPLOITING MORPHOLOGY IN SPEECH TRANSLATION WITH PHRASE-BASED FINITE-STATE TRANSDUCERS

*Alicia Pérez, M. Inés Torres**

Department of Electricity and Electronics
University of the Basque Country
manes.torres@ehu.es

Francisco Casacuberta

Instituto Tecnológico de Informática
Technical University of Valencia
fcn@iti.upv.es

ABSTRACT

This work implements a novel formulation for phrase-based translation models making use of morpheme-based translation units under a stochastic finite-state framework. This approach has an additional interest for speech translation tasks since it leads to the integration of the acoustic and translation models.

As a further contribution, this is the first paper addressing a Basque-to-Spanish speech translation task. For this purpose a morpheme based finite-state recognition system is combined with a finite-state transducer that translates phrases of morphemes in the source language into usual sequences of words in the target language.

The proposed models were assessed under a limited-domain application task. Good performances were obtained for the proposed phrase-based finite-state translation model using morphemes as translation units, and also notable improvements are obtained in decoding time.

Index Terms— Speech Translation, Stochastic Finite-State Transducers, Morphology

1. INTRODUCTION

The use of morphological knowledge in machine translation (MT) is relatively recent and has been mainly sustained in tasks where morphologically rich languages were involved. In both transfer-based and example-based MT approaches morphological analysis has been used in the source language to extract lemmas and split words into their compounds so as to predict word-forms in the target language [1, 2]. In [3] it was *Moses* [4], the state-of-the art statistical MT system, that was used to train phrase-based models at morpheme level.

With respect to MT under finite-state framework, in [5] a text-to-text translation paradigm was proposed by combining a phrase-based model dealing with running words and finite-state models including morphological knowledge. Specifically, the finite-state machine consisted of a composition of a word-to-stem statistical analyser in source word, a stem-to-stem translation model from source to target language and a stem-to-word statistical generation module in target language all the constituents being implemented with ATT-tools. No other morphemes except stems were used.

The contribution of this work is twofold: first, the formulation of speech translation based on morphemes under the finite-state framework, and second, its application on Basque to Spanish speech translation. We take advantage of all the compounds of a word, and not only of lemmas. We promote the use of finite-state models due to their decoding speed.

Spanish and Basque languages entail many challenges for current machine translation systems. Due to the fact that both languages are official in the Basque Country, there is a real demand of several documents to be bilingual. In spite of the fact that both languages coexist in the same area, they differ enormously. To begin with, it is precise to note that they have different origin: while Spanish belongs to the set of Romance languages, Basque is a pre-Indoeuropean language. There are notable differences in both morphology and syntax. In contrast to Spanish, Basque is an extremely inflected language, with more than 17 declension cases that can be recursively combined. Inflection makes the size of the vocabulary (in terms of word-forms) grow. Hence, the number of occurrences of word n-grams within the data is much smaller than in the case of Spanish, and this leads to poor or even unreliable statistic estimates. By applying to morpheme based models we aim at tackling sparsity of data and consequently getting improved statistical distributions.

2. MORPHEME-BASED SPEECH TRANSLATION

The goal of statistical speech translation is to find the most likely translation, $\hat{\bar{t}}$, given the acoustic representation, X, of a speech signal from the source language:

$$\hat{\bar{t}} = \arg\max_{\bar{t}} P(\bar{t}|X) \tag{1}$$

The transcription of speech in the source language into a sequence of morphemes, $\bar{m}$, can be introduced as a hidden vari-

*This work has been partially supported by the University of the Basque Country under grants 9/UPV 00224.310-15900/2004 and GIU07/57, by the Spanish CICYT under grant TIN2005-08660-C04-03, and by the Spanish program Consolider-Ingenio 2010 under grant CSD2007-00018.

able.

$$\hat{\hat{t}} = \arg\max_{\bar{t}} \sum_{\bar{m}} P(\bar{t}, \bar{m}|X) \quad (2)$$

Applying the Bayes' decision rule:

$$\hat{\hat{t}} = \arg\max_{\bar{t}} \sum_{\bar{m}} \frac{P(\bar{t}, \bar{m})P(X|\bar{t}, \bar{m})}{P(X)} \quad (3)$$

Let us assume that the probability of an utterance does not depend on the transcription in other language. Hence, the denominator would be independent of the variable over which the optimisation is being done, and thus, the decoding would be carried out as follows:

$$\hat{\hat{t}} = \arg\max_{\bar{t}} \sum_{\bar{m}} P(\bar{t}, \bar{m})P(X|\bar{m}) \quad (4)$$

It is the contribution of two terms that drives the search problem: 1) the acoustic model, $P(X|\bar{m})$, connecting a text string in terms of morphemes to its acoustic utterance; 2) the joint translation model, $P(\bar{t}, \bar{m})$, connecting source and target languages. Joint probability translation models are good candidates to be approached by stochastic finite-state transducers (SFSTs).

Some effort has been recently made in order to efficiently take advantage of both acoustic and translation knowledge sources [6] by exploring different architectures. We have implemented the morpheme-based speech translation models under two different architectures described in [7]: a) *integrated architecture* implementing eq. (4) analogously as in an automatic speech recognition (ASR) system where the LM was replaced by a joint probability model. Thanks to the nature of the finite state models a tight integration is allowed, making a difference with respect to other kind of integration; b) *decoupled architecture* where two stages are involved, that is, first, an ASR system copes with transcription of the speech utterance, and later, a text-to-text translation system translates the given transcription.

Finally, there is an important issue to be noted, and it is the fact that this formulation for speech translation makes use of morphemes only in the source language, while using word-forms in the target language. The underlying motivation is simply that a speech translation from a morphologically rich language into another that does not present inflection in nouns is being taken into consideration. This is, in fact, our case when translating from Basque to Spanish.

2.1. Phrase-based stochastic finite-state transducers

An SFST is a finite-state machine that analyses strings in a source language and accordingly produces strings in a target language along with the joint probability of both strings to be translation each other (for a formal definition turn to [6]). The characteristics defining the SFST are the topology and the probability distributions over the transitions and the states. These distinctive features can be automatically learnt from bilingual samples by efficient algorithms such as GIATI (Grammar Inference and Alignments for Transducers Inference) [7], which is applied in this work. As it is well known, an outstanding aspect of the finite-state models is the fact that they count on efficient standard decoding algorithms [8]. Indeed, it is the speed of the decoding stage that makes these models so attractive for speech translation.

In this work we deal with SFSTs based on phrases of morphemes. Previously, in [9], in phrase-based SFSTs were presented based on word-forms (we will refer to this approach as PW-SFST). In such a models the transitions occur consuming a sequence of words. Here we propose the use of sequences of morphemes PM-SFST instead. As for what the standard baseline SFST is concerned (referred to as W-SFST), the difference lies on the fact that the transitions consume isolated word-forms instead of sequences of either words or morphemes. In all the cases, the transitions of SFSTs produce a sequence of zero or more words in the target language and have a probability associated.

2.2. Morphological analysis

In this work we deal with a morphologically rich language: Basque. In Basque there is no freely available linguistic tool that splits the words into proper morphemes. For this reason, morpheme-like units were obtained by means of Morfessor [10], a data-driven approach based on unsupervised learning of morphological word segmentation. For both ASR and SMT it is convenient to keep a low morpheme to word ratio, in order to get better language modelling, acoustic separability and word generation amongst others. Consequently, in a previous work [11], an approach based of decomposing the words into two morpheme-like units, a *root* and an *ending* was presented. By default, Morfessor decomposed the words using 3 types of morphemes: prefixes, stems and suffixes. To convert the decompositions into the desired root-ending form, all the suffixes at the end of the word were joined to form the ending, and the root was built joining all the remaining prefixes, stems and possible suffixes between stems. This procedures led to a vocabulary of 946 morphemes set of [11].

3. EXPERIMENTAL RESULTS

Basque is a minority but official language in the Basque Country (Spain). It counts on scarce linguistic resources and database, in addition, it is a highly inflected language. As a result, exploiting the morphology seems a good choice to improve the reliance on statistics.

The models were assessed under METEUS corpus, consisting of a text and speech of weather forecast reports picked from those published in the Internet. As shown in Table 1, the corpus is divided into a training set and a training-independent test set consisting of 500 sentences. Each sentence of the test was uttered by at least 3 speakers, resulting in a speech evaluation data of 1,800 utterances from 36 speakers. Note that the size of the Basque vocabulary is 38% bigger than the Spanish one due to its inflected nature.

		Basque	Spanish
Training (Text)	Pair of sentences	14,615	
	Different pairs	8,220	
	Running words	154,778	168,722
	Vocabulary	1,097	675
	Average length	10.6	11.5
Test (Speech)	Utterances	1,800	
	Length (hours)	3.5	3.0

Table 1. Main features of METEUS corpus.

The phrase-based SFST using morphemes proposed here, PM-SFST, was compared with the other two models, previously mentioned, namely PW-SFST and W-SFST. The three models were trained from the corpus described in Table 1 making use of the so-called GIATI algorithm [7]. Speech translation was carried out using both the integrated and decoupled architectures. Besides, in order to explore the influence on the translation model of errors derived from the recognition process, a verbatim translation was also carried out. In this case, the input of the text-to-text translation system is the transcription of the speech free from errors (as if the recognition process had been flawless).

3.1. Computational cost and performance

The memory required for a model to be allocated in memory along with the invested decoding time are two key parameters to bear in mind when it comes to evaluating a speech translation system. Table 2 shows the spatial cost (in terms of number of transitions and branching factor) of each of the three SFST models studied along with the relative decoding time consumed. Regarding the time units, they are relative to the baseline W-SFST model, that is, given that the test was translated in 1 time unit by W-SFST, the time units required by the PW-SFST and PM-SFST was picked up.

	Transitions	BF	$<$Time$>$
W-SFST	114,531	3.27	1.00
PW-SFST	121,265	3.25	0.76
PM-SFST	127,312	3.21	0.71

Table 2. Spatial cost, in terms of number of transitions and branching factor (BF), and the relative amount of time required by each model for text-input translation (dimensionless magnitude).

Doubtless, it is the performance, measured in terms of translation accuracy or error rate what counts for the evaluation of both speech and text translation. Translation results were assessed under the commonly used automatic evaluation metrics: bilingual evaluation under study (BLEU [12]) and word error rate (WER). Table 3 shows speech translation results with the three approaches mentioned above and the different architectures. The recognition WER for decoupled architecture was obtained trough previous ASR experiments reported in [11] with the same set of moprhemes. We would like to emphasize that speech translation with integrated architecture gives both the transcription and the translation of speech in the same decoding step, as a result, and thus, each model gives its own recognition-word-error-rate.

		Recognition WER	Translation WER	Translation BLEU
Integrated	W-SFST	6.26	47.5	47.6
	PW-SFST	6.12	48.4	48.0
	PM-SFST	6.06	47.8	48.6
Decoupled	W-SFST	4.93	46.9	47.3
	PW-SFST	4.93	48.5	49.0
	PM-SFST	4.93	47.8	49.3
Verbatim	W-SFST	0	45.6	48.6
	PW-SFST	0	46.5	50.4
	PM-SFST	0	46.7	50.7

Table 3. Speech translation results provided by different translation models (W-SFST, PW-SFST, PM-SFST) under either integrated or decoupled architectures. The verbatim translation is also shown as a baseline.

3.2. Discussion

Both PM-SFST and PW-SFST models outperform the baseline W-SFST with 95% confidence under 1,000 bootstrap samples following the statistical significance test described in [13] with the BLEU evaluation measure. Nevertheless, the differences between PM-SFST and PW-SFST are marginal.

Comparing the two architectures considered, the translation results are similar. Furthermore, taking into account that the LM used for speech transcription in ASR with decoupled architecture and the SFST used to both recognize and translate speech counted on the same amount of data, one could expect that the parameters of the latter would not be as well considered, and accordingly, the performance of the integrated architecture would be worse for recognition purposes.

The differences in translation performance between speech translation with the decoupled architecture and the verbatim translation are small. There are two factors that have influence on this fact: on the one hand, the input of the speech translation was not very degraded; on the other hand, the transducer shows certain capacity to deal with input errors by mechanisms such as smoothing.

With respect to the size and time-efficiency of the models (summarized in Table 2), as it is obvious, the phrase-based models (both PM-SFST and PW-SFST) are bigger than W-SFST. Nevertheless, the branching factor is smaller, which indicates that the phrase-based models are more restrictive than the word-based in that, on average, they allow for a smaller number of transitions per state. Note that in the smoothed W-SFST all the strings have non-zero probability while in the phrase-based approaches only those strings built up in terms of the existing phrases have a non-zero probability. Regarding decoding time (in Table 2) there is a correlation with the branching factor. The higher the branching factor, the higher the required time, and thus, the PM-SFST model shows significant time reductions.

4. CONCLUDING REMARKS AND FUTURE WORK

For natural language processing applications when the language under study is morphologically rich, it might be useful to make use of morphology. By using morpheme-like units, statistics collected over a given database could be improved, and accordingly, the parameters describing statistical models. As far as speech translation is concerned, there is a further interest on the use of morphemes as lexical unit, and it is precisely that the way in which the morphemes were extracted kept a low morpheme to word ratio avoiding so acoustic confusion.

In this work we have dealt with Basque to Spanish speech translation. Morpheme-based speech translation has been proposed in terms of morphemes and within the finite-state framework. The models have been assessed under a limited-domain task giving as a result improvements in both translation accuracy and decoding time.

As far as future work is concerned, the generation of target words from morphemes given a source out of vocabulary word is still an open problem that might, as well, be explored from the statistical approach. That is, instead of doing analysing, as in our case, generation might be tackled.

5. REFERENCES

[1] G. Labaka, N. Stroppa, A. Way, and K. Sarasola, "Comparing rule-based and data-driven approaches to Spanish-to-Basque machine translation," in *Proc. Machine Translation Summit XI*, 2007.

[2] E. Minkov, K. Toutanova, and H. Suzuki, "Generating complex morphology for machine translation," in *Proc. 45st Annual Meeting of the Asocciation for Computational Linguistics*, 2007, pp. 128–135.

[3] S. Virpioja, J. J. Väyrynen, M. Creutz, and M. Sadeniemi, "Morphology-aware statistical machine translation based on morphs induced in an unsupervised manner," in *Proc. Machine Translation Summit XI*, 2007, pp. 491–498.

[4] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, et al., "Moses: Open source toolkit for statistical machine translation," *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion*, pp. 177–180, 2007.

[5] P. Karageorgakis, A. Potamianos, and I. Klasinas, "Towards incorporating language morphology into statistical machine translation systems," in *Proc. Automatic Speech Recogn. and Underst. Workshop (ASRU)*, 2005.

[6] F. Casacuberta, M. Federico, H. Ney, and E. Vidal, "Recent efforts in spoken language translation," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 80–88, 2008.

[7] F. Casacuberta and E. Vidal, "Learning finite-state models for machine translation," *Machine Learning*, vol. 66, no. 1, pp. 69–91, 2007.

[8] M. Mohri, F. Pereira, and M. Riley, "AT&T FSM LibraryTM and Finite-State Machine Library," 2003.

[9] A. Pérez, M. I. Torres, and F. Casacuberta, "Speech translation with phrase based stochastic finite-state transducers," in *Proc. IEEE 32nd International Conference on Acoustics, Speech, and Signal Processing* 2007, vol. IV, pp. 113–116, IEEE.

[10] M. Creutz and K. Lagus, "Inducing the morphological lexicon of a natural language from unannotated text," in *Proc. International and Interdisciplinary Conference on Aadaptive Knowledge Representation and Reasoning*, 2005.

[11] V. G. Guijarrubia, M. I. Torres, and R. Justo, "Morpheme-based automatic speech recognition of basque," in *Proc. 4th Iberian Conference on Pattern Recognition and Image Analysis*, 2009, pp. 386–393, Springer-Verlag.

[12] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proc. 40th Annual Meeting of the Association Computational Linguistics*, 2002, pp. 311–318.

[13] M. Bisani and H. Ney, "Bootstrap estimates for confidence intervals in ASR performance evaluation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004, vol. 1, pp. 409–412.