SPEECH TRANSLATION WITH PHRASE BASED STOCHASTIC FINITE-STATE TRANSDUCERS

*Alicia Pérez, M. Inés Torres**

Dep. Electricity and Electronics
University of the Basque Country
manes@we.lc.ehu.es

Francisco Casacuberta

Dep. Information Systems and Computation
Technical University of Valencia
fcn@dsic.upv.es

ABSTRACT

Stochastic finite-state transducers constitute a type of word-based models that allow an easy integration with acoustic model for speech translation. The aim of this work is to develop a novel approach to phrase-based statistical finite-state transducers. In this work, we explore the use of linguistically motivated phrases to build phrase-based models.

The proposed phrase-based transducer has been tested and compared to a word-based equivalent machine, yielding promising results in the reported preliminary text and speech translation experiments.

***Index Terms*—** Speech recognition, natural languages, transducers, statistics.

1. INTRODUCTION

Stochastic finite-state transducers (SFSTs) constitute a special type of word-based models that can be easily integrated with acoustic models for speech translation [1]. In this context, the goal of this work is to build phrase-based SFSTs.

Current trends in statistical machine translation aim beyond word-based models translation through alignment templates or phrase-based models (PBMs) that rely on clumps of statistically aligned sets of words [2]. It is well known that professional translators first split the source sentence into word-sequences (phrases) that play the same syntactic role to subsequently translate each phrase into the target language and arrange it according to the corresponding grammar rules. In that sense, statistical phrase-based translation models arise in an attempt to better mimic human behavior; and, in fact, they have proved to achieve better performance than word-based models [3].

One of the problems with PBMs is the selection of adequate bilingual phrases. There are different automatic techniques to select these units. The most usuals are example-based, with any kind of heuristics that mainly rely on word-to-word alignments [3, 4].These models make use of the word-level alignments provided by GIZA++ [5].

Given a parallel corpus GIZA++ aligns one source token to multiple target tokens. By default, space blanks are used to split the sentences into tokens, therefore, the running words are usually taken as tokens. In this work we are exploring the use of *linguistically motivated phrases* as tokens. Thus, we make sure that each token is going to be an *expression forming a grammatical constituent of a sentence*. Therefore, with this approach we are intending to align meanings instead of running words. One advantage of taking phrases as tokens is the reduction of the length of the sentences in each language. As a result, long-distance bilingual relationships can be cut down, and more accurate alignments can be obtained.

Another goal of this study is to apply, in a practical manner, the aforementioned model to Spanish-to-Basque translation, since this pair of languages show long distance alignments. Although Basque is a minority language, having only around 600,000 speakers, it enjoys official status along with Spanish in the Basque Country, hence, nowadays some resources are being devoted to machine translation, but they are still scarce. Even though they coexist in the same area, these languages differ greatly both in syntax as well as in semantics, which entails an added motivation for our work.

Most of the still little research on Basque is performed within linguistic framework, a dependency-based formalism defining the deep syntactic level of language description [6]. Some effort has been invested in the development of tagging and parsing. In this paper, we test the proposed phrase-based SFST over a Spanish-Basque application. The obtained results show a improvement over the word based SFST and also appropriate results for the studied pair of languages.

The organization of the article is as follows: in the next section we describe the stochastic finite-state transducers. Section 3 is devoted to the selection of linguistically motivated bilingual phrases. Experimental results are reported in Section 4. Finally, the conclusions that are drawn from this work are presented in the last section.

*This work has been partially supported by the Industry Department of the Basque Government and by the University of the Basque Country under grants INTEK CN02AD02 and 9/UPV 00224.310-15900/2004 respectively.

2. STOCHASTIC FINITE-STATE TRANSDUCERS

SFSTs constitute an interesting class of statistical translation models that have proved to be highly suitable for text-input and speech-input translation in specific tasks [7]. The definition (section 2.1) of these devices allow the use of phrases as input or output symbols, they are not restricted to running-words, and so are we going to proceed in this work.

2.1. Definition

An *stochastic finite-state transducer* (SFST) is a tuple $\mathcal{T} = \langle \Sigma, \Delta, Q, q_0, R, F, P \rangle$, where:

Σ is a finite set of input symbols (source vocabulary);

Δ is a finite set of output symbols (target vocabulary);

Q is a finite set of states;

$q_0 \in Q$ is the initial state;

$R \subseteq Q \times \Sigma \times \Delta^* \times Q$ is a set of transitions such as $(q, s, \tilde{t}, q')$, which is a transition from the state q to the state q', with the source symbol s and producing the substring $\tilde{t}$;

$P : R \rightarrow [0, 1]$ transition probability distribution;

$F : Q \rightarrow [0, 1]$ final state probability distribution;

The probability distributions satisfy the stochastic constraint:

$$\forall q \in Q \quad F(q) + \sum_{\forall s, \tilde{t}, q'} P(q, s, \tilde{t}, q') = 1 \tag{1}$$

2.2. Inference

Given a *bitext*, i.e. a set of training pairs consisting of sentences in the source and target languages, the structural and probabilistic components of the SFST can be automatically learned making use of a grammatical inference technique such as GIATI (*Grammar Inference and Alignments for Transducers Inference*) [7]. This method provides hybrid translation models since it combines the two main trends in statistical machine translation: statistical alignment models and finite state automaton models. GIATI is a general method without any practical restriction with regard to the used tokens, so we propose the use of this algorithm using phrases instead of running words as tokens in the training bitext.

One of the main problems related to PBM is the difficulty to generalize. That is, PBM models are less flexible than word-based models since the formers are less granular. Anyhow, finite-state models offer a natural method to deal with this problem, the so called *back-off* smoothing. When an input sentence can not be analyzed by the SFST through a direct path, it is possible to go back to a more general model, that is, from high n-gram models to lower ones, up to the unigram (if necessary) where all the words have been seen, even a special one, the "unknown" word, which is just a mechanism to model out of vocabulary words (OOV). Therefore, smoothed SFST models can cope with every single event in the test, including OOV, thus, they get the versatility required to deal with PBM.

2.3. Searching

SFSTs operate as follows: given an input-string ($\mathbf{s}$), the expected translation ($\widehat{\mathbf{t}}$) corresponds to the string that maximizes the joint probability:

$$\widehat{\mathbf{t}} = \arg\max_{\mathbf{t}} P_{\mathcal{T}}(\mathbf{s}, \mathbf{t}) \tag{2}$$

The probability associated to a pair $(\mathbf{s}, \mathbf{t})$, is the sum of the probabilities of all the possible paths that are compatible with that pair. Solving eq. (2) has proved to be a hard computational problem but it can be efficiently computed by *Viterbi algorithm* which obtains the best sequence of states through the SFST. The resulting translation is the result of the concatenation of the output strings through the optimal path. The searching engine implements a beam search strategy based on dynamic programming and allows for threshold pruning.

2.4. Integrated models for speech translation

A SFST can be easily integrated in a conventional speech recognition system (SRS) for speech-input machine translation purposes. As it is known, a SRS is composed by a language model and an acoustic model. The SRS makes use of the input projection of the SFST as if it was the common LM. As acoustic model, no further information other than the phonetic transcription of the input tokens in the SFST is needed. It does not matter what a token is, a running word or a phrase, we only care to have the appropriate transcription. In our system we make a composition of finite-state acoustic input strings in the SFST, therefore, a complete integration can be achieved in the framework of finite-state approach.

3. IDENTIFICATION OF LINGUISTIC PHRASES

The phrase identification is linguisticaly motivated and automatically run following the steps listed below:

1. First, a morpho-syntactic parse allows to assign either one or more tags to each word within the corpus. These tags include information about linguistic categories such as number, declension case, etc. Besides, the stem and the morphemes are identified. At this point, a word can be tagged with different categories, which means that it is an ambiguous word.

2. For each word, the most likely category is chosen bearing in mind that the word-sets in the sentence should share compatible categories.

		Translation unit			
		word		**phrase**	
		Spanish	Basque	Spanish	Basque
Training	Number of pairs	14,615			
	Different pairs	8,462			
	Different sentences	7,226	7,523	7,226	7,523
	Vocabulary	720	1,147	2,478	2,587
	Singeltons	178	294	970	111
	Mean length	13.0	12.8	7.1	9.4
Test-1	Number of pairs	1,500			
	Training independent	702			
	Different sentences	1,077	1,099	1,077	1,099
	OOV	0	0	52	69
	Mean length	12.6	12.4	6.9	9.1
	Perplexity (3-grams)	3.6	4.3	7.0	5.9
Test-2	Sentences	500			
	Training independent	500			
	Different sentences	500	500	500	500
	Mean length	17.4	16.5	9.5	12.0
	OOV	0	0	51	65
	Perplexity (3grams)	4.8	6.7	13.0	11.1

Table 1. Main features of the METEUS corpus when either the word or the linguistic-phrase are taken as token.

3. Once the syntactic and the semantic parsing of each element has been carried out unambiguously, linguistic phrases can be identified under a elementary criteria: to group, recursively, all the words that share the same syntactic function whenever the frequency of that cluster in the corpus exceeds a threshold.

This process entails a parser for each language under study, that is, the phrases obtained for each language are independent, they are selected without taking into account the other language. At this point the words are clustered in terms of meaning and taking the syntactic function as reference. The obtained phrases in one language usually have their counterpart in the other language, but obviously they do not appear in a monotone fashion.

A manual inspection shows that it is easier to find correspondences between bilingual phrases than between words, thus, our initial motivation is to provide GIZA++ with linguistic phrases instead of running words in order to obtain better quality alignments.

4. EXPERIMENTAL RESULTS

The METEUS corpus [8] consists of bilingual sentences extracted from weather forecast reports that had been published on the Internet by the Basque Institute of Meteorology. The main features of METEUS are shown in Table 1.

The SFST was learned from the training set described in Table 1 resorting to GIATI algorithm. Then, the obtained model was subjected to two different tests (Table 1). The first one consists of 1500 sentences randomly extracted from the whole corpus consists, thus, on average, it shows the same characteristics as the training corpus does. The second test set, was extracted from the former one in such a way that all the sentences were not included in the training set and all of them different each other. These two test sets would show us how the model works in two extreme situations, the most favorable case, for the first test, and the most pessimistic case, for the second one.

For speech input machine translation experiments, the second test set was recorded by 36 bilingual speakers uttering 50 sentence-pairs each, resulting in around 3.25 hours of audio signal for each language.

The 1500 test sentences were translated by the SFST, the total processing time being less than 60 seconds. The translation provided by the system was compared to the reference sentence (the one in the target language for that particular pair), and in order to quantitatively assess the performance of the system the following three evaluation parameters were defined:

WER: *Word Error Rate* is the relative number of error edit operations between the reference sentence and the output of the system (the lower the better).

PER: *Position independent Error Rate* is similar to the WER but without taking into account the word ordering within the sentence.

BLEU: *BiLingual Evaluation Understudy* is based on the n-grams of the hypothesized translation that occur in the reference translations. The BLEU metric ranges from 0.0 (worst score) to 1.0 (best score).

Translation results, provided by both the word-based and phrase-based SFST devices, are shown in Table 2. All the SFST models have been built under the same 3-gram approach. As can be inferred from the data, better translations are obtained with the phrase-based transducer in all the circumstances.

The word-based SFST has got 148,367 states whereas phrase based SFST has got 61,100 states, that is, less than a half. Therefore, with regard to spatial and temporal costs, once again, phrase-based models outperform word-based ones.

The weak spots of finite-state transducers are, first, the difficulty to keep relationships between input and output languages at a long distance, and second, the amount of memory they consume. Both of them have been overcome through the development of the phrase-based approach to finite-state devices proposed here, since distances are cut down and less amount of states are needed when taking the phrase as translation unit. And that is fulfilled even for the language pair under consideration in the present work, consisting of two very different languages where long distance correspondences are frequent. Spanish into Basque translation has been proved to be a difficult task comparing to Spanish into English task under the similar conditions.

	test-1-T		test-2-T		test-2-S	
	WB	**PB**	**WB**	**PB**	**WB**	**PB**
PER	31.4	23.3	41.6	37.7	55.6	47.0
WER	36.6	28.6	50.1	47.5	57.6	56.5
BLEU	0.56	0.71	0.42	0.46	0.37	0.44

Table 2. Translation performance with word-based (WB) and phrase-based (PB) 3-gram SFST models for different tests: test-1 with text-input (test-1-T), test-2 with text-input (test-2-T) and test-2 with speech-input (test-2-S).

5. CONCLUDING REMARKS

This paper describes an extension to a word based FST, where translation units are linguistically determined phrases. Phrase-based SFST have been developed and tested. The nature of the phrases taken into account in this work are linguistical. Taking word-based models as a baseline, phrase-based models yield a better performance, both in terms of error rates and of spatial cost. Therefore, phrases seem to be a good choice for finite-state transducers, and also for the language pairs taken into account.

Future work could attempt at including more linguistic knowledge in the transducer or different ways to obtain the phrases.

6. ACKNOWLEDGMENTS

We would like to thank *Ametzagaiña* group[1], and Josu Landa, in particular, for providing us with the morpho-syntactic parse which made possible this work.

7. REFERENCES

[1] F. Casacuberta, H. Ney, F. J. Och, E. Vidal, J. M. Vilar, S. Barrachina, I. García-Varea, D. Llorens, C. Martínez, S. Molau, F. Nevado, M. Pastor, D. Picó, A. Sanchis, and C. Tillmann, "Some approaches to statistical and finite-state speech-to-speech translation," *Computer Speech and Language*, vol. 18, pp. 25–47, Jan. 2004.

[2] F.J. Och and H. Ney, "The alignment template approach to statistical machine translation," *Computational Linguistics,*, vol. 30, no. 4, pp. 417–450, 2004.

[3] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation," in *Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference (HLT/NAACL)*, Edmonton, Canada, May 2003.

[4] J. Tomás and F. Casacuberta, "Monotone statistical translation using word groups," in *Proceedings of the Machine Translation Summit VIII*, Santiago de Compostela, Spain, September 2001, pp. 357–361.

[5] Franz Josef Och and Hermann Ney, "Statistical machine translation," in *Proc. of Workshop of the European Association for Machine Translation*, Ljubljana, Slovenia, May 2000, pp. 39–46.

[6] Antonio M. Corbí-Bellot, Mikel L. Forcada, Sergio Ortiz-Rojas, Juan Antonio Pérez-Ortiz, Gema Ramírez-Sánchez, Felipe Sánchez-Martínez, Iñaki Alegria, Aingeru Mayor, and Kepa Sarasola, "An open-source shallow-transfer machine translation engine for the romance languages of spain," in *Proceedings of the Tenth Conference of the European Associtation for Machine Translation*, Budapest, Hungary, May 2005, pp. 79–86.

[7] F. Casacuberta and E. Vidal, "Machine translation with inferred stochastic finite-state transducers," *Computational Linguistics*, vol. 30, no. 2, pp. 205–225, 2004.

[8] A. Pérez, M. Inés Torres, F. Casacuberta, and V. Guijarrubia, "A Spanish-Basque weather forecast corpus for probabilistic speech translation," in *Proceedings of the 5th SALTMIL Workshop on Minority Languages*, Genoa, Italy, 2006.

[1] `http://www.ametza.com`