Speech-input multi-target machine translation

Alicia Pérez, M. Inés Torres
Dep. of Electricity and Electronics
University of the Basque Country
manes@we.lc.ehu.es

M. Teresa González, Francisco Casacuberta
Dep. of Information Systems and Computation
Technical University of Valencia
fcn@dsic.upv.es

Abstract

In order to simultaneously translate speech into multiple languages an extension of stochastic finite-state transducers is proposed. In this approach the speech translation model consists of a single network where acoustic models (in the input) and the multilingual model (in the output) are embedded.

The multi-target model has been evaluated in a practical situation, and the results have been compared with those obtained using several mono-target models. Experimental results show that the multi-target one requires less amount of memory. In addition, a single decoding is enough to get the speech translated into multiple languages.

1 Introduction

In this work we deal with finite-state models which constitute an important framework in syntactic pattern recognition for language and speech processing applications (Mohri et al., 2002; Pereira and Riley, 1997). One of their outstanding characteristics is the availability of efficient algorithms for both optimization and decoding purposes.

Specifically, stochastic finite-state transducers (SFSTs) have proved to be useful for machine translation tasks within restricted domains. There are several approaches implemented over SFSTs which range from word-based systems (Knight and Al-Onaizan, 1998) to phrase-based systems (Pérez et al., 2007). SFSTs usually offer high speed during the decoding step and they provide competitive results in terms of error rates. In addition, SFSTs have proved to be versatile models, which can be easily integrated with other finite-state models, such as a speech recognition system for speech-input translation purposes (Vidal, 1997). In fact, the integrated architecture has proved to work better than the decoupled one. Our main goal is, hence, to extend and assess these methodologies to accomplish spoken language multi-target translation.

As far as multilingual translation is concerned, there are two main trends in machine translation devoted to translate an input string simultaneously into m languages (Hutchins and Somers, 1992): *interlingua* and *parallel transfer*. The former has historically been a knowledge-based technique that requires a deep-analysis effort, and the latter consists on m decoupled translators in a parallel architecture. These translators can be either knowledge or example-based. On the other hand, in (González and Casacuberta, 2006) an example based technique consisting of a single SFST that cope with multiple target languages was presented. In that approach, when translating an input sentence, only one search through the multi-target SFST is required, instead of the m independent decoding processes required by the mono-target translators.

The classical layout for speech-input multi-target translation includes a speech recognition system in a serial architecture with m decoupled text-to-text translators. Thus, this architecture entails a decoding stage of the speech signal into the source language text, and m further decoding stages to translate the source text into each of the m target lan-

Proceedings of the Second Workshop on Statistical Machine Translation, pages 56–63,
Prague, June 2007.

guages. If we supplant the m translators with the multi-target SFST, the problem would be reduced to 2 searching stages. Nevertheless, in this paper we propose a natural way for acoustic models to be integrated in the multilingual network itself, in such a way that the input speech signal can be simultaneously decoded and translated into m target languages. As a result, due to the fact that there is just a single searching stage, this novel approach entails less computational cost.

The remainder of the present paper is structured as follows: section 2 describes both multi-target SFSTs and the inference algorithm from training examples; in section 3 a novel integrated architecture for speech-input multi-target translation is proposed; section 4 presents a practical application of these methods, including the experimental setup and the results they produced; finally, section 5 summarizes the main conclusions of this work.

2 Multi-target stochastic finite-state transducers

A multi-target SFST is a generalization of standard SFSTs, in such a way that every input string in the source language results in a tuple of output strings each being associated to a different target language.

2.1 Definition

A *multi-target stochastic finite-state transducer* is a tuple $\mathcal{T} = \langle \Sigma, \Delta_1 \ldots \Delta_m, Q, q_0, R, F, P \rangle$, where:

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

$\Delta_1 \ldots \Delta_m$ are m finite sets of output symbols (target vocabularies);

Q is a finite set of states;

$q_0 \in Q$ is the initial state;

$R \subseteq Q \times \Sigma \times \Delta_1^* \ldots \Delta_m^* \times Q$ is a set of transitions such as $(q, w, \tilde{p}_1, \ldots, \tilde{p}_m, q')$, which is a transition from the state q to the state q', with the source symbol w and producing the substrings $(\tilde{p}_1, \ldots, \tilde{p}_m)$;

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

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

The probability distributions satisfy the stochastic constraint:

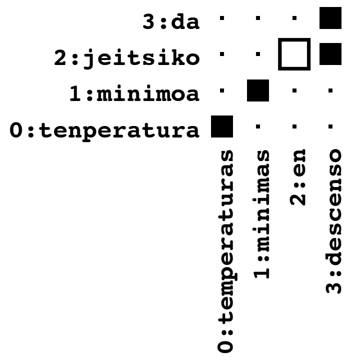
$$\forall q \in Q \qquad (1)$$
$$F(q) + \sum_{w, \tilde{p}_1, \ldots, \tilde{p}_m, q'} P(q, w, \tilde{p}_1, \ldots, \tilde{p}_m, q') = 1$$

2.2 Training the multilingual translation model

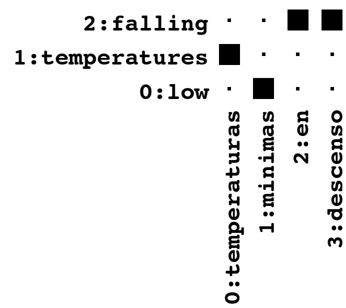
Both topology and parameters of an SFST can be learned fully automatically from bilingual examples making use of underlying alignment models (Casacuberta and Vidal, 2004). Furthermore, a multi-target SFST can be inferred from a multilingual set of samples (González and Casacuberta, 2006). Even though in realistic situations multilingual corpora are too scarce, recent works (Popović et al., 2005) show that bilingual corpora covering the same domain are sufficient to obtain generalized corpora based on which one can subsequently create the required collections of aligned tuples.

The inference algorithm, GIAMTI (*grammatical inference and alignments for multi-target transducer inference*), requires a multilingual corpus, that is, a finite set of multilingual samples $(\mathbf{s}, \mathbf{t_1}, \ldots, \mathbf{t_m}) \in \Sigma^* \times \Delta_1^* \times \cdots \times \Delta_m^*$, where $\mathbf{t_i}$ denotes the translation of the source sentence $\mathbf{s}$ into the i-th target language; Σ denotes the source language vocabulary, and Δ_i the i-th target language vocabulary; the algorithm can be outlined as follows:

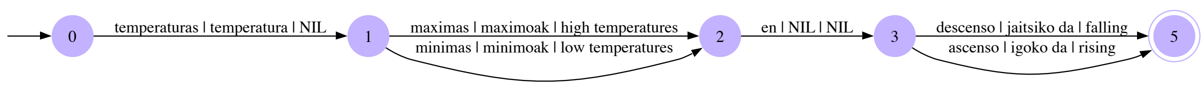
1. Each multilingual sample is transformed into a single string from an *extended vocabulary* ($\Gamma \subseteq \Sigma \times \Delta_1^* \times \cdots \times \Delta_m^*$) using a *labeling function* ($\mathcal{L}^m$). This transformation searches an adequate monotonic segmentation for each of the m source-target language pairs on the basis of bilingual alignments such as those given by GIZA++ (Och, 2000). A monotonic segmentation copes with monotonic alignments, that is, $j < k \Rightarrow a_j < a_k$ following the notation of (Brown et al., 1993). Each source token, which can be either a word or a phrase (Pérez et al., 2007), is then joined with a target phrase of each language as the corresponding segmentation suggests. Each *extended symbol* consists of a token from the source language plus zero



(a) *Spanish-Basque*



(b) *Spanish-English*



(c) *Multi-target SFST from Spanish into English and Basque.*

Figure 1: *Example of a trilingual alignment over a trilingual sentence extracted from the task under consideration;the related multi-target SFST (with Spanish as input, and English and Basque as output).*

or more words from each target language in their turn.

2. Once the set of multilingual samples has been converted into a set of single extended strings ($\mathbf{z} \in \Gamma^*$), a stochastic regular grammar can be inferred. Specifically, in this work we deal with k-testable in the string-sense grammars (García and Vidal, 1990), which are considered to be a syntactic approach of the n-gram models. In addition, they allow the integration of several order models in a single smoothed automaton (Torres and Varona, 2001).

3. The extended symbols associated with the transitions of the automaton are transformed into one input token and m output phrases ($w/\tilde{p}_1|\ldots|\tilde{p}_m$) by the inverse labeling function ($\mathcal{L}^{-m}$), leading to the required transducer.

Example An illustration of the inference of the multi-target SFST can be shown over a couple of simple trilingual sentences from the corpus (where "B" stands for Basque, "S" for Spanish and "E" for English):

1-B tenperatura maximoa jaitsiko da

1-S temperaturas máximas en descenso

1-E high temperatures falling

2-B tenperatura minimoa igoko da

2-S temperaturas mínimas en ascenso

2-E low temperatures rising

From the alignments, depicted in Figures 1(a) and 1(b), an input-language-synchronized monotonous segmentation can be built (bear in mind that we are considering Spanish as the input language). The corresponding extended strings with the following constituents for the first and second samples respectively are the following ones:

1 temperaturas|tenperatura|λ
mínimas|minimoa|low_temperatures
en|λ|λ
descenso|jaitsiko_da|falling

2 temperaturas|tenperatura|λ
máximas|maximoa|high_temperatures
en|λ|λ
ascenso|igoko_da|rising

Finally, from this representation of the data, the multi-target SFST can be built as shown in Figure 1(c).

2.3 Decoding

Given an input string $\mathbf{s}$ (a sentence in the source language), the decoding module has to search the optimal m output strings $\mathbf{t}^m \in \Delta_1^* \times \cdots \times \Delta_m^*$ (a sentence in each of the target language) according to the underlying translation model ($\mathcal{T}$):

$$\widehat{\mathbf{t}^m} = \arg \max_{\mathbf{t}^m \in \Delta_1^* \times \cdots \times \Delta_m^*} P_{\mathcal{T}}(\mathbf{s}, \mathbf{t}^m) \quad (2)$$

Solving equation (2) is a hard computational problem, however, it can be efficiently computed under the so called *maximum approach* as follows:

$$P_{\mathcal{T}}(\mathbf{s}, \mathbf{t}^m) \approx \max_{\phi(\mathbf{s}, \mathbf{t}^m)} P_{\mathcal{T}}(\phi(\mathbf{s}, \mathbf{t}^m)) \quad (3)$$

where $\phi(\mathbf{s}, \mathbf{t}^m)$ is a *translation form*, that is, a sequence of transitions in the multi-target SFST compatible with both the input and the m output strings.

$$\phi(\mathbf{s}, \mathbf{t}^m) : (q_0, w_1, \tilde{p}_1^m, q_1) \cdots (q_{J-1}, w_J, \tilde{p}_J^m, q_J)$$

The input string ($\mathbf{s}$) is a sequence of J input symbols, $\mathbf{s} = w_1^J$, and each of the m output strings consists of J phrases in its corresponding language $\mathbf{t}^m = (\mathbf{t}_1, \cdots, \mathbf{t}_m) = (\tilde{p_1})_1^J, \cdots, (\tilde{p_m})_1^J$. Thus, the probability supplied by the multi-target SFST to the translation form is given by:

$$P_{\mathcal{T}}(\phi(\mathbf{s}, \mathbf{t}^m)) = F(q_J) \prod_{j=1}^{J} P(q_{j-1}, w_j, \tilde{p}_j^m, q_j) \quad (4)$$

In this context, the *Viterbi algorithm* can be used to obtain the optimal sequence of states through the multi-target SFST for a given input string. As a result, the established m translations are built concatenating the (J) output phrases for each language through the optimal path.

3 An embedded architecture for speech-input multi-target translation

3.1 Statistical framework

Given the acoustic representation ($\mathbf{x}$) of a speech signal, the goal of multi-target speech translation is to find the most likely m target strings ($\mathbf{t}^m$); that is, one string ($\mathbf{t}_i$) per target language involved ($i \in \{1, \ldots, m\}$). This approach is summarized in eq. (5), where the hidden variable $\mathbf{s}$ can be interpreted as the transcription of the speech signal:

$$\widehat{\mathbf{t}^m} = \arg \max_{\mathbf{t}^m} P(\mathbf{t}^m | \mathbf{x}) = \arg \max_{\mathbf{t}^m} \sum_{\mathbf{s}} P(\mathbf{t}^m, \mathbf{s} | \mathbf{x}) \quad (5)$$

Making use of Bayes' rule, the former expression turns into:

$$\widehat{\mathbf{t}^m} = \arg \max_{\mathbf{t}^m} \sum_{s} P(\mathbf{t}^m, \mathbf{s}) P(\mathbf{x} | \mathbf{t}^m, \mathbf{s}) \quad (6)$$

Empirically, there is no loss of generality if we assume that the acoustic signal representation depends only on the source string, i.e. $P(\mathbf{x}|\mathbf{t}^m, \mathbf{s})$ is independent of $\mathbf{t}^m$. In this sense, eq. (6) can be rewritten as:

$$\widehat{\mathbf{t}^m} = \arg \max_{\mathbf{t}^m} \sum_{s} P(\mathbf{t}^m, \mathbf{s}) P(\mathbf{x} | \mathbf{s}) \quad (7)$$

Equation (7) combines a standard acoustic model, $P(\mathbf{x}|\mathbf{s})$, and a multi-target translation model, $P(\mathbf{t}^m, \mathbf{s})$, both of whom can be integrated on the fly during the searching routine as shown in Figure 2. That is, each acoustic sub-network is only expanded at decoding time when it is required.

The outer sum is computationally very expensive to search for the optimal tuple of target strings $\mathbf{t}^m$ in an effective way. Thus we make use of the so called Viterbi approximation, which finds the best path over the whole transducer.

3.2 Practical issues

The underlying recognizer used in this work is our own continuous-speech recognition system, which implements stochastic finite-state models at all levels: acoustic-phonetic, lexical and syntactic, and which allows to infer them based on samples.

The signal analysis was carried out in a standard way, based on the classical Mel-cepstrum parametrization. Each phone-like unit was modeled

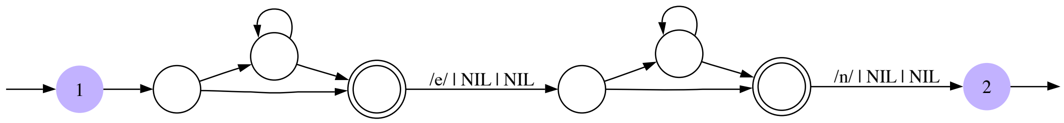


Figure 2: *Integration on the fly of acoustic models in one edge of the SFST shown in Figure 1(c)*

by a typical left to right hidden Markov model. A phonetically-balanced Spanish database, called Albayzin (Moreno et al., 1993), was used to train these models.

The lexical model consisted of the extended tokens of the multi-target SFST instead of running words. The acoustic transcription for each extended token was automatically obtained on the basis of the input projection of each unit, that is, the Spanish vocabulary in this case.

Instead of the usual language model, we make use of the multi-target SFST itself, which had the syntactic structure provided by a k-testable in the strict sense model, with k=3, and Witten-Bell smoothing. Note that the SFST implicitly involves both input and output language models.

4 Experimental results

4.1 Task and corpus

The described general methodology has been put into practice in a highly practical application that aims to translate on-line TV weather forecasts into several languages, taking the speech of the presenter as the input and producing as output text-strings, or sub-titles, in several languages. For this purpose, we used the corpus METEUS which consists of a set of trilingual sentences, in English, Spanish and Basque, as extracted from weather forecast reports that had been published on the Internet. Let us notice that it is a real trilingual corpus, which they are usually quite scarce.

Basque is a pre-Indoeuropean language of still unknown origin. It is a minority language, spoken in a small area of Europe and also within some small American communities (such as that in Reno, Nevada). In the Basque Country (located in the north of Spain) it has an official status along with Spanish. However, despite having coexisted for centuries in the same area, they differ greatly both in syntax and in semantics. Hence, efforts are being devoted nowadays to machine translation tools involving these two languages (Alegria et al., 2004), although they are still scarce. With regard to the order of the phrases within a sentence, the most common one in Basque is *Subject plus Objects plus Verb* (even though some alternative structures are also accepted), whereas in Spanish and English other constructions such as *Subject plus Verb plus Objects* are more frequent (see Figures 1(a) and 1(b)). Another difference between Basque and Spanish or English is that Basque is an extremely inflected language.

In this experiment we intend to translate Spanish speech simultaneously into both Basque and English. Just by having a look at the main features of the corpus in Table 1, we can realize that there are substantial differences among these three languages, in terms both of the size of the vocabulary and of the amount of running words. These figures reveal the agglutinant nature of the Basque language in comparison with English or Spanish.

		Spanish	**Basque**	**English**
Training	Total sentences	14,615		
	Different sentences	7,225	7,523	6,634
	Words	191,156	187,462	195,627
	Vocabulary	702	1,147	498
	Average Length	13.0	12.8	13.3
Test	Sentences	500		
	Words	8,706	8,274	9,150
	Average Length	17.4	16.5	18.3
	Perplexity (3grams)	4.8	6.7	5.8

Table 1: Main features of the METEUS corpus.

With regard to the speech test, the input consisted of the speech signal recorded by 36 speakers, each one reading out 50 sentences from the test-set in Table 1. That is, each sentence was read out by at least three speakers. The input speech resulted in approximately 3.50 hours of audio signal. Needless to say, the application that we envisage has to be speakerindependent if it is to be realistic.

4.2 System evaluation

The performance obtained by the acoustic integration has been experimentally tested for both multi-target and mono-target devices. As a matter of comparison, text-input translation results are also reported.

The multi-target SFST was learned from the training set described in Table 1 using the previously described GIAMTI algorithm. The 500 test sentences were then translated by the multi-target SFST. The translation provided by the system in each language was compared to the corresponding reference sentence. Additionally, two mono-target SFSTs were inferred with their outputs for the aforementioned test to be taken as baseline. The evaluation includes both computational cost and performance of the system.

4.2.1 Computational cost

The expected searching time and the amount of memory that needs to be allocated for a given model are two key parameters to bear in mind in speech-input machine translation applications. These values can be objectively measured in terms of the size and on the average branching factor of the model displayed in Table 2.

	multi-target	**mono-target**	
		S2B	**S2E**
Nodes	52,074	35,034	20,148
Edges	163,146	115,526	69,690
Branching factor	3.30	3.13	3.46

Table 2: Features of multi-target model and the two decoupled mono-target models (one for Spanish to Basque translation, referred to as S2B, and the second for Spanish to English, S2E).

Adding the edges up for the two mono-target SFSTs that take part in the decoupled architecture (see Table 2), we conclude that the decoupled model needs a total of $185,216$ edges to be allocated in memory, which represents an increment of 13% in memory-space with respect to the multi-target model.

On the other hand, the multi-target approach offers a slightly smaller branching factor than each mono-target approach. As a result, fewer paths have to be explored with the multi-target approach than with the decoupled one, which suggests that searching for a translation might be faster. As a matter of fact, experimental results in Table 3 show that the mono-target architecture works 11% more slowly than the multi-target one for speech-input machine translation and decoding, and 30% for text to text translation.

	Time (s)	
	multi-target	**mono-target S2B+S2E**
Text-input	0.36	0.47
Speech-input	16.9	18.9

Table 3: Average time needed to translate each input sentence into two languages.

Summarizing, in terms of computational cost (space and time), a multi-target SFST performs better than the mono-target decoupled system.

4.2.2 Performance

So far, the capability of the systems has been assessed in terms of time and spatial costs. However, the quality of the translations they provide is, doubtless, the most relevant evaluation criterion. In order to determine the performance of the system in a quantitative manner, the following evaluation parameters were computed for each scenario: *bilingual evaluation under study* (BLEU), *position independent error rate* (PER) and *word error rate* (WER). Both text and speech-input translation results provided by the multi-target and the mono-target models respectively are shown in Table 4.

As can be derived from the translation results, for text-input translation the classical approach performs slightly better than the multi-target one, but for speech-input translation from Spanish into English is the other way around. In any case, the differences in performance are marginal.

Comparing the text-input with the speech-input results we realize that, as could be expected, the process of speech signal decoding is itself introducing some errors. In an attempt to measure these errors, the text transcription of the recognized input signal was extracted and compared to the input reference in terms of WER as shown in the last row of the Table 4. Note that even though the input sentences are the same the three results differ due to the fact that we are making use of different SFST models that decode and translate at the same time.

		multi-target		mono-target	
		S2B	**S2E**	**S2B**	**S2E**
Text	**BLEU**	42.7	66.7	43.4	67.8
	PER	39.9	19.9	38.2	19.0
	WER	48.0	27.5	46.2	26.6
Speech	**BLEU**	39.5	59.0	39.2	61.1
	PER	42.2	25.3	41.5	23.6
	WER	51.5	33.9	50.5	31.9
	recognition WER	10.7		9.3	9.1

Table 4: Text-input and speech-input translation results for Spanish into Basque (S2B) and Spanish into English (S2E) using a multi-target SFST (columns on the left) or two mono-target SFSTs (columns on the right). The last row shows Spanish speech decoding results using each of the three devices.

In these series of experiments the same task has been compared with two extremely different language pairs under the same conditions. There is a noticeable difference in terms of quality between the English and the Basque translations. The underlying reason might be due to the fact that SFST models do not capture properly the rich morphology of the Basque as they have to face long-distance reordering issues. These differences in the performance of the system when translating into English or into Basque have been previously detected in other works (Ortiz et al., 2003). In our case, a manual review of the models and the obtained translations encourage us to make use of reordering models in future work, since they have proved to report good results in a similar framework (Kanthak et al., 2005).

5 Concluding remarks and further work

The main contribution of this paper is the proposal of a fully embedded architecture for multiple speech translation. Thus, acoustic models are integrated on the fly into a multi-target translation model. The most significant feature of this approach is its ability to carry out both the recognition and the translation into multiple languages integrated in a unique model. Due to the finite-state nature of this model, the speech translation engine is based on a Viterbi-like algorithm.

In contrast to the mono-target systems, multi-target SFSTs enable the translation from one source language simultaneously into several target languages with lower computational costs (in terms of space and time) and comparable qualitative results. Moreover, the integration of several languages and acoustic models is straightforward on means of finite-state devices.

Nevertheless, the integrated architecture needs more parameters to be estimated. In fact, as the amount of targets increase the data sparseness might become a difficult problem to cope with. In future work we intend to make a deeper study on the performance of the multi-target system with regard to the amount of parameters to be estimated. In addition, as the first step of the learning algorithm is decisive, we are planning to make use of reordering models in an attempt to face up to with long distance reordering and in order to homogenize all the languages involved.

Acknowledgments

This work has been partially supported by the University of the Basque Country and by Spanish CICYT under grants 9/UPV 00224.310-15900/2004, TIC2003-08681-C02-02, and CICYT es TIN2005-08660-C04-03 respectively.

References

Iñaki Alegria, Olatz Ansa, Xabier Artola, Nerea Ezeiza, Koldo Gojenola, and Ruben Urizar. 2004. Representation and treatment of multiword expressions in basque. In Takaaki Tanaka, Aline Villavicencio, Francis Bond, and Anna Korhonen, editors, *Second ACL Workshop on Multiword Expressions: Integrating Processing*, pages 48–55, Barcelona, Spain, July. Association for Computational Linguistics.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and R. L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

Francisco Casacuberta and Enrique Vidal. 2004. Machine translation with inferred stochastic finite-state transducers. *Computational Linguistics*, 30(2):205–225.

P. García and E. Vidal. 1990. Inference of k-testable languages in the strict sense and application to syntactic pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):920–925.

M.T. González and F. Casacuberta. 2006. Multi-Target Machine Translation using Finite-State Transducers. In *Proceedings of TC-Star Speech to Speech Translation Workshop*, pages 105–110.

John Hutchins and Harold L. Somers. 1992. *An Introduction to Machine Translation*. Academic Press, Cambridge, MA.

Stephan Kanthak, David Vilar, Evgeny Matusov, Richard Zens, and Hermann Ney. 2005. Novel reordering approaches in phrase-based statistical machine translation. In *Proceedings of the ACL Workshop on Building and Using Parallel Texts*, pages 167–174, Ann Arbor, Michigan, June. Association for Computational Linguistics.

K. Knight and Y. Al-Onaizan. 1998. Translation with finite-state devices. In *4th AMTA (Association for Machine Translation in the Americas)*.

Mehryar Mohri, Fernando Pereira, and Michael Riley. 2002. Weighted finite-state transducers in speech recognition. *Computer, Speech and Language*, 16(1):69–88, January.

A. Moreno, D. Poch, A. Bonafonte, E. Lleida, J. Llisterri, J. B. Mario, and C. Nadeu. 1993. Albayzin speech database: Design of the phonetic corpus. In *Proc. of the European Conference on Speech Communications and Technology (EUROSPEECH)*, Berlín, Germany.

Franz J. Och. 2000. GIZA++: Training of statistical translation models. `http://www.fjoch.com/GIZA++.html`.

Daniel Ortiz, Ismael García-Varea, Francisco Casacuberta, Antonio Lagarda, and Jorge González. 2003. On the use of statistical machine translation techniques within a memory-based translation system (AMETRA). In *Proc. of Machine Translation Summit IX*, pages 115–120, New Orleans, USA, September.

Fernando C.N. Pereira and Michael D. Riley. 1997. Speech Recognition by Composition of Weighted Finite Automata. In Emmanuel Roche and Yves Schabes, editors, *Finite-State Language Processing*, Language, Speech and Communication series, pages 431–453. The MIT Press, Cambridge, Massachusetts.

Alicia Pérez, M. Inés Torres, and Francisco Casacuberta. 2007. Speech translation with phrase based stochastic finite-state transducers. In *Proceedings of the 32nd International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, Honolulu, Hawaii USA, April 15-20. IEEE.

Maja Popović, David Vilar, Hermann Ney, Slobodan Jovičić, and Zoran Šarić. 2005. Augmenting a small parallel text with morpho-syntactic language. In *Proceedings of the ACL Workshop on Building and Using Parallel Texts*, pages 41–48, Ann Arbor, Michigan, June. Association for Computational Linguistics.

M. Inés Torres and Amparo Varona. 2001. k-tss language models in speech recognition systems. *Computer Speech and Language*, 15(2):127–149.

Enrique Vidal. 1997. Finite-state speech-to-speech translation. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 111–114, Munich, Germany, April.