Natural Language Engineering **2** (4) 351–354. Printed in the United Kingdom

© 1996 Cambridge University Press

Text and speech translation by means of subsequential transducers

J. M. VILAR,[1] V. M. JIMÉNEZ,[2] J. C. AMENGUAL,[1]
A. CASTELLANOS,[1] D. LLORENS[2] and E. VIDAL[2]

[1]*Unidad Predepart. de Informática, Campus de Penyeta Roja*
Universitat Jaume I, E-12071 Castellón de la Plana (SPAIN)
[2]*Depto. de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia*
Camino de Vera s/n, E-46071 Valencia (SPAIN)
e-mail: jvilar@inf.uji.es

(*Received 31 December 1996; Revised 6 February 1997*)

The full paper explores the possibility of using Subsequential Transducers (SST), a finite state model, in limited domain translation tasks, both for text and speech input. A distinctive advantage of SSTs is that they can be efficiently learned from sets of input-output examples by means of OSTIA, the Onward Subsequential Transducer Inference Algorithm (Oncina *et al.* 1993). In this work a technique is proposed to increase the performance of OSTIA by reducing the asynchrony between the input and output sentences, the use of error correcting parsing to increase the robustness of the models is explored, and an integrated architecture for speech input translation by means of SSTs is described.

A SST (Berstel 1979) is a *deterministic* finite state machine that accepts sentences from a given input language and produces associated sentences of an output language. Despite the conceptual simplicity of SSTs, the results show that they can perform surprisingly well in limited domain tasks. A SST implicitly defines a deterministic finite state automaton, which determines the set of strings for which there is a translation by the SST. This automaton is referred to as the *input language model* of the SST and it can be extended to asign probabilities to the input language sentences. The parameters of this stochastic extension can be easily estimated by maximum likelihood.

There are cases in which not only a good translation model but also a good input language model is required. This happens when distorted input (for instance the result of OCR or speech recognition modules) has to be translated. Although the SSTs learned by OSTIA are usually good translation models, they do not define good input language models, due to overgeneralization during learning. An extension of OSTIA, known as OSTIA-DR (Oncina and Varó 1996), allows us to control this problem by imposing syntactic constraints to the input (domain) and output (range) languages of the learned SST. In particular, the domain and range models used with OSTIA-DR can be n-testable automata, which are equivalent to n-grams.

https://doi.org/10.1017/S1351324997001617 Published online by Cambridge University Press

Table 1. *Translation WER (on an independent test set) and size of the transducers obtained with the original and reordered corpora*

Training size	Original			Reordered		
	WER	States	Edges	WER	States	Edges
1000	55.1	813	2023	31.7	532	1338
2000	47.1	1406	3353	10.3	358	979
4000	30.1	1686	4051	1.4	144	440
8000	1.4	244	719	0.5	109	344
16000	0.0	100	363	0.0	63	183

An additional and important problem is the excessive growth of the number of states of the SST (and of the required number training pairs) due to differences in the word order between the input and output languages. This 'asynchrony' can be reduced by changing the order of the output sentences in the training corpus so that they follow more closely the input order, leaving marks so that the correct ordering can be recovered later, and using OSTIA-DR to learn a transducer with this new corpus. This transducer can then be used to translate new sentences, obtaining output sentences that will need to be reordered to get the correct translation.

If a partial alignment between the input and output training sentences is known, a reordering can be easily obtained (Vilar *et al.* 1996). An alignment can be obtained in different ways. For instance, a dictionary can be used in order to find the correspondences between the input and output words. A dictionary based on the results of IBM's 'Model 1' (Brown *et al.* 1993) was used in the experiments reported here. From the resulting dictionary, function words were eliminated and those word pairs with a high probability were used to define a partial alignment.

A sentence may fail to be translated by an SST if it does not comply with the syntactic restriction of the SST, either because of errors in the sentence (e.g. if it was obtained by OCR) or lack of training material for the SST. In both cases, Error-Correcting Parsing (ECP) techniques can be used to increase the robustness of SSTs (Amengual *et al.* 1996; Vilar *et al.* 1996). Under the ECP approach, the input sentence is considered as a *corrupted* version (according to some error model) of some correct sentence in the input language of the SST. Parsing therefore consists in finding this sentence in the input language, which can be efficiently done for error models that comprise insertions, substitutions and deletions using fast ECP techniques (Amengual *et al.* 1996). Therefore combined ECP and translation can be performed to find both a best error corrected sentence and its translation.

Speech input language translation by means of SSTs can be approached in two ways: in a *decoupled* manner the SSTs learned by OSTIA would translate the output of a CSR system; in an *integrated approach* the input language model of a conventional CSR system (Ney 1995) is replaced by a SST, which is used both as language and translation model in a simultaneous search for the recognised sentence and its corresponding translation (Jiménez *et al.* 1995).

The performance of the SSTs with text and speech input has been tested with

https://doi.org/10.1017/S1351324997001617 Published online by Cambridge University Press

Table 2. *Effect of ECP for three types of test input sentences:* clean, uniformly distorted *with* 5% *distortion, and* distorted *to* 50% *using a dictionary of* spontaneous language *effects. (The models were trained with reordered output.)*

Training size	Clean Test Set		5%–Distorted Test Set		50%–Spont. Test Set	
	No ECP	ECP	No ECP	ECP	No ECP	ECP
1000	31.7	16.4	67.7	17.6	99.5	16.8
2000	10.3	5.0	59.0	6.2	99.4	4.8
4000	1.4	0.6	55.0	2.2	99.4	0.6
8000	0.5	0.1	54.6	1.7	99.3	0.1
16000	0.0	0.0	54.4	1.7	99.3	0.1

Table 3. *Results for speech input, with and without reordering*

Training size	Recognition WER		Translation WER	
	Original	Reordered	Original	Reordered
1000	5.9	4.0	45.5	17.5
2000	5.0	3.6	38.7	7.3
4000	4.3	3.1	28.2	3.2
8000	3.9	2.9	4.8	3.0
16000	3.0	3.2	3.2	3.3
32000	3.5	3.2	3.9	3.3
50000	3.5	3.2	3.7	3.3

the Miniature Language Acquisition (MLA) task (Feldman *et al.* 1990) extended and adapted to be a Spanish-to-English language translation task (Castellanos *et al.* 1994). A first series of experiments with *clean text* were carried out to determine the impact of reordering in the models learned. Both the translation *Word Error Rate* (WER) and the evolution of model sizes clearly assess the superior performance of models learned with reordering (see Table 1). The next series of experiments was aimed at assessing the capabilities of ECP both to improve the performance of insufficiently trained models (clean test set), and to recover from 'distorted' input data (see Table 2). Error and language model probabilities were estimated using versions of the training data that were also distorted as explained in (Amengual *et al.* 1996). The last series of experiments assessed the performance of SSTs with speech input. The results in Table 3 show the usefulness of the word reordering technique to reduce the number of training pairs required to obtain a given error rate in an integrated speech input translation system. Previous experiments (not reported here) demonstrated both the advantages of an integrated approach over a decoupled approach for speech translation, and that the use of n-gram input language models to constrain the SST learned by OSTIA-DR leads to translation models whose input language model is better that the original n-gram model (Jiménez *et al.* 1995).

From all these experiments, we can conclude that the use of SSTs is adequate for limited domain tasks, ECP techniques can be employed to improve their robustness,

https://doi.org/10.1017/S1351324997001617 Published online by Cambridge University Press

and they can be integrated with standard finite state lexical and acoustic models in order to translate speech input. The building of all these models can be done at a low cost since all of them can be automatically learned from training data.

Clearly, the bottleneck of this approach is finding large corpora of paired sentences in different languages. It is likely that these resources will be available in the very near future, in the same way as corpora for training acoustic and language models exist today. Techniques like the reordering of the output sentences and the ECP techniques presented here can also be used to reduce the amount of data needed.

References

Amengual, J. C., Vidal, E., and Benedí, J. M. 1996. Simplifying Language through Error-Correcting Decoding. Proceedings of the ICSLP-96, pp. 841–844.

Berstel, J. 1979. *Transductions and Context-Free Languages*. Stuttgart: Teubner.

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., and Mercer, R. L. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics* **19**: 263–311.

Castellanos, A., Galiano, I., and Vidal, E. 1994. Application of OSTIA to Machine Translation Tasks. In R. C. Carrasco and J. Oncina (eds.), *Grammatical Inference and Applications*, pp. 93–105. LNAI 862. Springer-Verlag.

Feldman, J. A., Lakoff, G., Stolcke, A., and Weber, S. H. 1990. Miniature Language Acquisition: A Touchstone for Cognitive Science. Technical Report TR-90-009, ICSI, Berkeley, California.

Jiménez, V. M., Castellanos, A., and Vidal, E. 1995. Some Results with a Trainable Speech Translation and Understanding System. Proceedings of the ICASSP-95, pp. 113-116.

Ney, H. 1995. Search strategies for large-vocabulary continuous-speech recognition. In A. J. Rubio and J. M. López (eds.), *Speech Recognition and Coding, New Advances and Trends*, pp. 210–225. NATO Advanced Study Institute. Springer-Verlag.

Oncina, J., García, P., and Vidal, E. 1993. Learning Subsequential Transducers for Pattern Recognition Interpretation Tasks. *IEEE Transactions on PAMI* **15**: 448–458.

Oncina, J., and Varó, M. A. 1996. Using Domain Information During the Learning of a Subsequential Transducer. In L. Miclet and C. De La Higuera (eds.), *Grammatical Inference: Learning Syntax from Sentences*, pp. 301–312. LNAI 1147. Berlin: Springer-Verlag.

Vilar, J. M., Vidal, E. and Amengual, J.C. 1996. Learning Extended Finite State Models for Language Translation. Proceedings of the ECAI-96, pp. 92–96.

https://doi.org/10.1017/S1351324997001617 Published online by Cambridge University Press