Factored Translation Models for improving a Speech into Sign Language Translation System

V. López-Ludeña, R. San-Segundo, R. Córdoba, J. Ferreiros, J.M. Montero, J.M. Pardo

Grupo de Tecnología del Habla.
Universidad Politécnica de Madrid
{veronicalopez|lapiz|cordoba|jfl|juancho|pardo}@die.upm.es

Abstract

This paper proposes the use of Factored Translation Models (FTMs) for improving a Speech into Sign Language Translation System. These FTMs allow incorporating syntactic-semantic information during the translation process. This new information permits to reduce significantly the translation error rate. This paper also analyses different alternatives for dealing with the non-relevant words. The speech into sign language translation system has been developed and evaluated in a specific application domain: the renewal of Identity Documents and Driver's License. The translation system uses a phrase-based translation system (Moses). The evaluation results reveal that the BLEU (BiLingual Evaluation Understudy) has improved from 69.1% to 73.9% and the mSER (multiple references Sign Error Rate) has been reduced from 30.6% to 24.8%.

Index Terms: Factored Translation Models, Speech into Sign Language translation, Phrase-based translation model.

1. Introduction

Based on information from INE (Spanish Statistics Institute http://www.ine.es) and the MEC (Ministry of Education www.educacion.es) 92% of deaf people in Spain have significant difficulties in understanding and expressing themselves in written Spanish. The main problems are related to verb conjugations, gender/number concordances and abstract concepts explanations. In 2007, the Spanish Government accepted Spanish Sign Language (LSE: Lengua de Signos Española) as one of the official languages in Spain, defining a long-term plan to invest in new resources for developing, disseminating and increasing the standardization of this language.

One important problem is that LSE is not disseminated enough among hearing people. This is why there are communication barriers between deaf and hearing people. These barriers are even more problematic when they appear between a deaf person and a government employee who is providing a face-to-face service, since they can cause the deaf people to have fewer opportunities. This happens, for example, when people want to renew the Identity Document or the Driver's License (DL). Generally, a lot of government employees do not know LSE so a deaf person needs an interpreter for accessing these services.

This paper proposes to use Factored Translation Models (FTMs) for improving the performance of a Speech into Sign Language Translation System. This system helps deaf people to communicate with government employees in a restricted domain: the renewal of Identity Documents and Driver's License [1]. This system has been designed to translate the government employee's explanations into LSE when government employees provide these personal services. The system is made up of a speech recognizer (for decoding the spoken utterance into a word sequence), a natural language translator (a phrase-based system for converting a word sequence into a sequence of signs belonging to the sign language), and a 3D avatar animation module (for playing back the signs) (Figure 1).

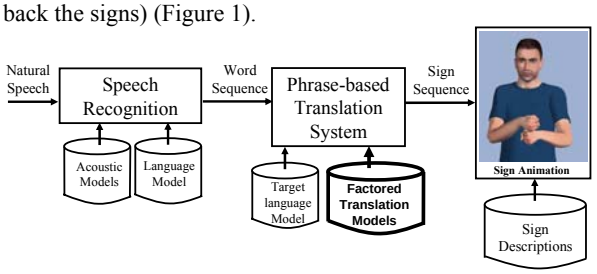

Figure 1. Spanish into LSE translation system.

2. Background

In the last ten years, there have been several efforts in order to develop prototypes for translating Spoken language into Sign Language: example-based [2], rule-based [1], full sentence [3] or statistical [4][5][6] approaches. None of these systems have analyzed the effect of incorporating additional information by using FTMs as it is proposed in this paper.

The main problem when researching in translation methods involving sign languages is the sparseness of data. Because of this, in the last 5 years, several projects have started to generate more resources. One of the most ambitious one is focused on generating a corpus made up of more than 300 hours from 100 speakers in Australian Sign Language [7]. The RWTH-BOSTON-400 Database contains 843 sentences with about 400 different signs from 5 speakers in American Sign Language with English annotations [8]. The British Sign Language Corpus Project aims to create a machine-readable digital corpus of spontaneous and elicited British Sign Language (BSL) collected from deaf native signers and early learners across the United Kingdom [9]. There are others examples in ISL (Irish Sign Language) [10], NGS (German Sign Language) [11], GSK (Greek Sign Language) [12] and Italian Sign Language [13].

For LSE, the biggest database was generated two years ago in a Plan Avanza project (www.traduccionvozlse.es) [14] and it is has been used in this work. Not only the data but also new practice [15] and new uses of traditional annotation tools [16] have been developed.

3. Database

The database used in this work was obtained with the collaboration of Local Government Offices where the renewal of Identity Document and Driver's license services are provided. For a period of three weeks, the most frequent explanations (from government employees) and the most frequent questions (from the user) were taken down. Finally, 4,080 sentences were collected [14]. These sentences were translated into LSE, both in text (sequence of signs) and in video, and compiled in an Excel file. Translation was carried out by two LSE experts in parallel. An example of Spanish sentence would be "en esta hoja viene todo lo necesario (this document contains all you need)" and his LSE translation "PAPEL ESTE INFORMACIÓN DETALLE TODO (DOCUMENT THIS INFORMATION DETAIL ALL)". The main features of the corpus are summarized in Table 1.

	Spanish	LSE
Sentence pairs	4,080	
Different sentences	3,342	1,289
Words or signs per sentence	7.7	5.7
Running words	31,501	23,256
Vocabulary	1,232	636

Table 1. Main statistics of the corpus

The corpus was divided randomly into three sets: training (75%), development (12.5%) and test (12.5%). The training set was used for tuning the speech recognizer (vocabulary and language model) and training the translation models. The development set was used for tuning the phrase-based translation system and finally, the test set was used for evaluating the approach proposed in this paper.

4. Automatic Speech Recognizer

The Automatic Speech Recognizer (ASR) used is a state-of-the-art speech recognition system developed at GTH-UPM (http://lorien.die.upm.es). It is a speaker independent continuous speech recognition system based on HMMs (Hidden Markov Models). The feature extraction includes CMN and CVN (Cepstrum Medium and Variance Normalization) techniques. The ASR provides one confidence value for each word recognized in the word sequence. Regarding the performance of the ASR module, with vocabularies smaller than 1,000 words, the Word Error Rate (WER) is lower than 5%.

5. Phrase-based translation system

The Phrase-based translation system is based on the software released to support the shared task at the 2010 NAACL Workshop on Statistical Machine Translation (http://www.statmt.org/wmt09/) [17]. Figure 2 shows the system architecture.

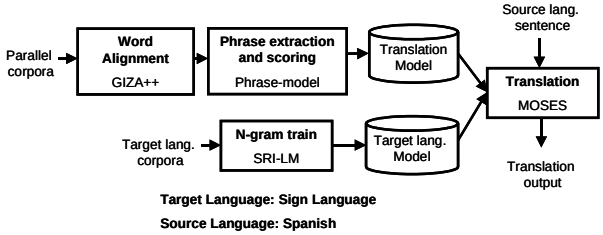

Figure 2. Diagram of the phrase-based translation module

The phrase model has been trained following three main steps.

- The first one is the word alignment computation using GIZA++. GIZA++ is a statistical machine translation toolkit that is used to train IBM Models 1-5 and an HMM (Hidden Markov Model) word alignment model. This package also contains the source for the mkcls tool which generates the word classes necessary for training some of the alignment models. In this case, GIZA++ has been used to calculate the alignments between words and glosses in both direction (Spanish-LSE and LSE-Spanish). For generating translation model, the parameter "alignment" was fixed to "target-source" as the best option: only this target-source alignment is considered (LSE-Spanish). In this configuration, alignment is guided by signs: each sign in LSE is aligned with a Spanish word and it is possible that some words were not aligned to any sign.
- The second step is phrase extraction. All phrase pairs that are consistent with the word alignment (target-source alignment in our case) are collected. The maximum size of a phrase has been fixed to 7 based on tuning experiments with the development set.
- The last step is phrase scoring. In this step, the translation probabilities are computed for all phrase pairs. Both translation probabilities are calculated: forward and backward.

The Moses decoder is used for the translation process. This program is a beam search decoder for phrase-based statistical machine translation models. In order to obtain a 3-gram language model, the SRI language modeling toolkit has been used [18].

6. Factored Translation Models

Considering a phrase-based translation strategy, there is the possibility to train factored models in order to include additional information in the translation process [19]. This possibility is an extension of phrase-based statistical machine translation models that enables the straightforward integration of additional annotations at the word-level (linguistic markup or automatically generated word classes). The main idea is to add additional annotation at the word level. A word in this framework is not only a token, but a vector of factors that represents different levels of annotation. The translation of factored representations of input words into the factored representations of output words is broken up into a sequence of mapping steps that either translates input factors into output factors, or generates additional output factors from existing output factors.

The information included in these factored models can be a tag with semantic information, sort of word (name, article, verb, adverb, preposition, etc.), gender or number of word, verb tense, adverb characteristics, etc. So, words in corpus become as a vector with the next format: Word|Factor1|Factor2|... For example, word "documento" becomes:"documento|DOCUMENTACIÓN|nombre|singular"

This paper proposes to add a new factor with synthetic-semantic information in the source language (Spanish). For adding this new factor, the categorization module used in the rule-based translation system previously developed for this application domain [1] has been considered. This rule-based translation system is composed of two main modules. In the first one, every word is mapped into one syntactic-pragmatic tag (categorization module). After that, the translation module applies different rules that convert the tagged words

into signs by means of grouping concepts or signs and defining new signs.

In order to use this categorization module, three different strategies was considered for dealing with "non-relevant" words, words that are not relevant for the translation process. They are tagged with the non-relevant tag named "basura" (garbage).

In the first alternative, all the words in the source language are factored and several translations models are trained (word-sign and tag-sign). Only two factors have been considered: word and tag. This alternative will be referred in the experiments like "**Using tags**". For example:

- Source sentence: debes pagar las tasas en la caja (you must pay the taxes in the cash desk)
- Factorized source sentence: debes|DEBER pagar|PAGAR las|basura tasas|DINERO en|basura la|basura caja|DINERO=CAJA
- Target sentence: VENTANILLA ESPECÍFICO CAJA TU PAGAR (WINDOW SPECIFIC CASH YOU PAY)

The second proposed alternative was to keep the original words (without additional factors), but removing non-relevant words from the source lexicon. This alternative will be referred in the experiments like "**Removing non-relevant words from the source lexicon**".

- Source sentence: debes pagar las tasas en la caja (you must pay the taxes in the cash desk)
- Factorized source sentence: debes pagar tasas caja
- Target sentence: VENTANILLA ESPECÍFICO CAJA TU PAGAR (WINDOW SPECIFIC CASH YOU PAY)

Finally, in third alternative all the words are factored and "non-relevant" words are removed. This alternative will be referred in the experiments like "**Using tags and removing non-relevant tags**".

- Source sentence: debes pagar las tasas en la caja (you must pay the taxes in the cash desk)
- Factorized source sentence: debes|DEBER pagar|PAGAR tasas|DINERO caja|DINERO=CAJA
- Target sentence: VENTANILLA ESPECÍFICO CAJA TU PAGAR (WINDOW SPECIFIC CASH YOU PAY)

7. Experiments and discussion

For the experiments, the corpus (described in section 3) was divided randomly into three sets: training (75%), development (12.5%) and test (12.5%). Results are compared with a baseline. This baseline consists of training models with original source and target corpus without any type of factorization, i.e, sentences contains words and signs from the original database. For example: this sentence "debes pagar las tasas en la caja" (you must pay the taxes in the cash desk) is translated into "VENTANILLA ESPECÍFICO CAJA TU PAGAR" (WINDOW SPECIFIC CASH YOU PAY).

For evaluating the performance of the translation systems, different accuracy metrics are presented: BLEU (BiLingual Evaluation Understudy) [20] in percentage and NIST [21]. Both metrics are computed using the NIST tool (mteval.pl). Additionally, two error metrics have been also added to the results: mSER (multiple references Sign Error Rate) and PER (multiple reference Position independent sign Error Rate). It is important to notice that BLEU and NIST are accuracy metrics while mSER and PER are error metrics. In order to analyze the significance of the differences between several systems, for every BLEU result, the confidence interval (at 95%) is also presented. This interval is calculated using the following formula:

$$\pm \Delta = 1{,}96\sqrt{\frac{BLEU(100-BLEU)}{n}} \qquad (1)$$

n is the number of signs used in evaluation, in this case n=2,906.

Table 2 compares the baseline system and the system with the FTMs for translating the references (Reference) and the speech recognizer outputs (ASR output).When using the FTMs the three different alternatives for dealing with non-relevant words are analyzed. Comparing these three alternatives, it is shown that adding tags to the words and removing "non-relevant" words are complementary actions that allow reaching the best results. When analyzing errors produced by the system, there are three main types of errors:

1. One of the most important types of error is related to the fact that in Spanish there are more words than signs in LSE (7.7 for Spanish and 5.7 for LSE in this corpus). This circumstance provokes the generation of many phrases in the same output: producing a high number of insertions. Additionally, when dealing with long sentences there is the risk that the translation model can not deal properly with the big distortion. This distortion produces important changes in order and sometimes the sentence is truncated producing several deletions.
2. Secondly, when translating Spanish into LSE, there is a relevant number of words in the testing set that they do not appear in the training set due to the higher variability presented in Spanish. For example, verb conjugations. In Spanish there are many verb conjugations that are translated into the same sign sequence. So, when a new conjugation appears in the evaluation set, it provokes a translation error.

Translation system		BLEU(%)	±Δ	NIST	mSER(%)	PER(%)
Baseline	Reference	73.7	1.6	8.6	26.9	19.5
	ASR output	**69.1**	1.7	**8.0**	**30.6**	**24.1**
Using tags	Reference	75.5	1.6	8.6	26.6	20.7
	ASR output	**68.0**	1.7	**7.8**	**31.5**	**26.4**
Removing non-relevant words from the source lexicon	Reference	80.0	1.4	8.9	20.4	17.1
	ASR output	**73.9**	1.6	**8.3**	**25.1**	**21.7**
Using tags and removing "non-relevant" tags	Reference	81.8	1.4	9.0	19.4	17.1
	ASR output	**73.9**	1.6	**8.3**	**24.8**	**22.7**

Table 2. Evaluation results including the baseline system and the system with the FTMs. In both cases, using these systems to translate the references (Reference) and the speech recognizer outputs (ASR output).

3. Finally, other important source of errors corresponds to ordering errors provoked by the different order in predication: LSE has a SOV (Subject-Object-Verb) while Spanish SVO (Subject-Verb-Object).

In conclusion, the main causes of the translation errors are related to the different variability in the vocabulary for Spanish and LSE (much higher in Spanish), the different number for words or signs in the sentences (higher in Spanish) and the different predication order.

The FTMs, including synthetic-semantic information, allow reducing the variability in the source language (for example, several verb conjugations are tagged with the same tag) and also the number of tokens composing the input sentence (when removing non-relevant words). These two aspects allow increasing the system performance significantly.

Also, reducing the source language variability and the number of tokens provoke an important reduction on the number of source-target alignments the system has to train. When having a small corpus, as it is the case of many sign languages, this reduction of alignment points permits to obtain better training models with less data, improving the results. BLEU has increased from 73.7% to 81.8% when translating reference sentences and from 69.1% to 73.9% when translating ASR outputs.

8. Conclusions

This paper describes the use of Factored Translation Models (FTMs) for improving a phrase-based Speech into Sign Language Translation System. This system is used to translate government employee's explanations into LSE when providing a personal service for renewing the Identity Document and Driver's License. These FTMs allow incorporating syntactic-semantic information during the translation process. This information reduces the variability in the source language and also the number of tokens composing the input sentence. These two aspects permit to increase the translation system performance.

9. Acknowledgements

The authors would like to thank the eSIGN consortium for permitting the use of the eSIGN Editor and the 3D avatar. The authors would also like to thank discussions and suggestions from the colleagues at GTH-UPM. This work has been supported by Plan Avanza Exp Nº: TSI-020100-2010-489), INAPRA (MEC ref: DPI2010-21247-C02-02), and SD-TEAM (MEC ref: TIN2008-06856-C05-03) projects and FEDER program.

10. References

[1] San-Segundo R., Barra R., Córdoba R., D'Haro L.F., Fernández F., Ferreiros J., Lucas J.M., Macías-Guarasa J., Montero J.M., Pardo J.M, 2008. "Speech to Sign Language translation system for Spanish". Speech Communication, Vol 50. 1009-1020. 2008.

[2] Morrissey, S. 2008. "Data-Driven Machine Translation for Sign Languages". Thesis. Dublin City University, Dublin, Ireland.

[3] Cox, S.J., Lincoln M., Tryggvason J., Nakisa M., Wells M., Mand Tutt, and Abbott, S., 2002 "TESSA, a system to aid communication with deaf people". In ASSETS 2002, pages 205-212, Edinburgh, Scotland, 2002.

[4] Stein, D., Bungeroth, J. and Ney, H.: 2006 "Morpho-Syntax Based Statistical Methods for Sign Language Translation". 11th Annual conference of the European Association for Machine Translation, Oslo, Norway, June 2006.

[5] Morrissey S., Way A., Stein D., Bungeroth J., and Ney H., 2007 "Towards a Hybrid Data-Driven MT System for Sign Languages. Machine Translation Summit (MT Summit)", pages 329-335, Copenhagen, Denmark, September 2007.

[6] Vendrame M., Tiotto G., 2010. ATLAS Project: Forecast in Italian Sign Language and Annotation of Corpora. In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010.

[7] Johnston T., 2008. "Corpus linguistics and signed languages: no lemmata, no corpus". 3rd Workshop on the Representation and Processing of Sign Languages, June 1. 2008.

[8] Dreuw P., Neidle C., Athitsos V., Sclaroff S., and Ney H. 2008. "Benchmark Databases for Video-Based Automatic Sign Language Recognition". In International Conference on Language Resources and Evaluation (LREC), Marrakech, Morocco, May 2008.

[9] Schembri. A.,2008 "British Sign Language Corpus Project: Open Access Archives and the Observer's Paradox". Deafness Cognition and Language Research Centre, University College London. LREC 2008.

[10] Morrissey S., Somers H., Smith R., Gilchrist S., Dandapat S., 2010 "Building Sign Language Corpora for Use in Machine Translation". In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010.

[11] Hanke T., König L., Wagner S., Matthes S.. 2010. "DGS Corpus & Dicta-Sign: The Hamburg Studio Setup". In 4th Workshop on the Representation and Processing of Sign Languages (CSLT 2010), Valletta, Malta, May 2010.

[12] Efthimiou E., and Fotinea, E., 2008 "GSLC: Creation and Annotation of a Greek Sign Language Corpus for HCI" LREC.

[13] Geraci C., Bayley R., Branchini C., Cardinaletti A., Cecchetto C., Donati C., Giudice S., Mereghetti E., Poletti F., Santoro M., Zucchi S. 2010. "Building a corpus for Italian Sign Language. Methodological issues and some preliminary results". In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010.

[14] San-Segundo, R., Pardo, J.M., Ferreiros, F., Sama, V., Barra-Chicote, R., Lucas, JM., Sánchez, D., García. A., "Spoken Spanish Generation from Sign Language" Interacting with Computers, Vol. 22, No 2, pp. 123-139, 2010.

[15] Forster J., Stein D., Ormel E., Crasborn O., Ney H.. 2010. "Best Practice for Sign Language Data Collections Regarding the Needs of Data-Driven Recognition and Translation". In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010.

[16] Crasborn O., Sloetjes H.. 2010. "Using ELAN for annotating sign language corpora in a team setting". In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010.

[17] Koehn, Philipp. 2010. "Statistical Machine Translation". Cambridge University Press.

[18] Stolcke A., 2002. "SRILM – An Extensible Language Modelling Toolkit". Proc. Intl. Conf. on Spoken Language Processing, vol. 2, pp. 901-904, Denver.

[19] Koehn, P., Hoang, H., "Factored Translation Models". 2007. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 868–876, Prague, June 2007.

[20] Papineni K., S. Roukos, T. Ward, W.J. Zhu. 2002 "BLEU: a method for automatic evaluation of machine translation". 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, PA, pp. 311-318. 2002.

[21] Doddington, G. 2002 "Automatic evaluation of machine translation quality using n-gram cooccurrence statistics". Proceedings of the Human Language Technology Conference (HLT), San Diego, CA pp. 128–132.