# Spanish generation from Spanish Sign Language using a phrase-based translation system

*Verónica López, Rubén San-Segundo, Raquel Martín, Juan Manuel Lucas, Julián David Echeverry*

Speech Technology Group at Universidad Politécnica de Madrid
veronicalopez@die.upm.es

## Abstract

This paper describes the development of a Spoken Spanish generator from Spanish Sign Language (LSE – Lengua de Signos Española) in a specific domain: the renewal of Identity Document and Driver's license. The system is composed of three modules. The first one is an interface where a deaf person can specify a sign sequence in sign-writing. The second one is a language translator for converting the sign sequence into a word sequence. Finally, the last module is a text to speech converter. Also, the paper describes the generation of a parallel corpus for the system development composed of more than 4,000 Spanish sentences and their LSE translations in the application domain. The paper is focused on the translation module that uses a statistical strategy with a phrase-based translation model, and this paper analyses the effect of the alignment configuration used during the process of word-based translation model generation. Finally, the best configuration gives a 3.90% mWER and a 0.9645 BLEU.

**Index Terms**: Automatic Statistic Translation, Sign Language Translation, Spoken Spanish Generation, Phrase-based Translator, Type of Word Alignment.

## 1. Introduction

There are approximately 70 million people with hearing deficiencies in the world (information from World Federation of the Deaf). Deafness brings about significant problems in communicating, because deaf people cannot hear and most of them are unable to use written languages, having big problems in understanding and expressing this way (they have problems with verb tenses, concordances of gender and number, etc., and they have difficulties when creating a mental image of abstract concepts). This fact can cause deaf people to have problems in accessing to information, education, job, social relationship, culture, etc., because they use a sign language for communicating and there is not a sufficient number of sign-language interpreters and communication systems for it.

In the same way, deaf people have problems when they want to access to public services, for example, to renew their Driver's License (DL) or Identity Document (ID). In general, government employees do not know LSE, so a deaf person needs an interpreter for accessing to this service.

In 2007, the Spanish Government accepted the Spanish Sign Language (LSE: Lengua de Signos Española) as one of the official languages in Spain, defining a plan to invest in resources in this language and that it becomes not only the natural language for deaf people, but also an instrument when communicating with hearing people, or accessing information. The translation system described in this paper is part of this government plan and its goal is to help deaf people to communicate with government employees in two specific domains: the renewal of ID and DL.

## 2. Spanish Sign Language

Spanish Sign Language (LSE), just like other sign languages, has a visual-gestural channel, but it also has grammatical characteristics of oral languages that shares with exclusive characteristics of sign languages. In linguistic terms, sign languages are as complex as oral languages, despite the common misconception that they are a "simplification" of oral languages. The main characteristics of LSE and the differences with Spanish are as follows [11]:

- Predication order: LSE has a SOV (subject-object-verb) order in opposite to SVO (subject-verb-object) Spanish order. For example:
  **Spanish**: *Juan ha comprado las entradas*
  **LSE**: *JUAN ENTRADAS COMPRAR*
- Gender in LSE is not usually specified.
- For specifying verb tenses, the verb tense can be added in parentheses next to the gloss, for example "*USAR (FUT.) (to use in future)*".
- For representing a negative sentence, it is added to the verb in infinitive, the gloss "*NO*", for example "*PODER NO*" (cannot).
- In LSE, also there are spelling for representing names or unknown words and this is indicated with "dl" previous to the spelled word, for example, "*dlJUAN*" for spelling the name "*Juan*".
- The use of classifiers is common in LSE: signs that indicate actions, places, etc., and that are denoted with the prefix "*CL*" and a letter that indicates the classifier's type (for example, place). Some classifiers are "*CLL-ACERCARSE*", "*CLD-GRANDE*", etc. There are not classifiers in Spanish.
- Spanish has an informative style (without topics) and LSE has a communicative style (with topics).
- In LSE, there can be concordances between verbs and subject, receiver or object and even subject and receiver, but in Spanish there can be only concordance between verb and subject.
- Articles are used in Spanish, but not in LSE
- Plural can be descriptive in LSE, but not in Spanish.
- There is a difference between absent and present third person in LSE, but there is not absent third person in Spanish.
- In LSE, there is the possibility of using double reference, not in Spanish.
- LSE is a language with ample flexibility, and homonymy between substantive and adjective is usual, so most nouns can be adjectives and vice versa. But there are few cases in Spanish.
- In Spanish, there is a copula in non-verbal predications (the verb "to be", *ser* and *estar* in Spanish), but there is not in LSE (except some locative predications).

- There is a difference between inclusive and exclusive quantifier in LSE, but not in Spanish.
- There are Spanish impersonal sentences with "*se*" pronoun, but not in LSE.
- Iconicity: signs resemble to concept that represent. If written LSE is analysed, glosses have semantic information principally.
- It is important to comment that LSE is more lexically flexible than Spanish, and it is perfect for generating periphrasis through its descriptive nature and because of this, LSE has fewer nouns than Spanish.
- LSE has less gloss per sentence (4.4) than Spanish (5.9).

# 3. State of the Art

Several groups have generated corpora for sign language research. Some examples are: a corpus composed of more than 300 hours from 100 speakers in Australian Sign Language [1]. The RWTH-BOSTON-400 Database that contains 843 sentences with about 400 different signs from 5 speakers in American Sign Language with English annotations [2]. The British Sign Language Corpus Project tries 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 [3]. And a corpus developed at Institute for Language and Speech Processing (ILSP) and that contains parts of free signing narration, as well as a considerable amount of grouped signed phrases and sentence level utterances [4].
The best performing translation systems are based on various types of statistical approaches ([5]; [6]), including example-based methods [7], finite-state transducers [8] and other data driven approaches. Another important effort in machine translation has been the organization of several Workshops on Statistical Machine Translation (SMT). As a result of these workshops, there are two free machine translation systems called Moses (http://www.statmt.org/moses/) and Joshua (http://cs.jhu.edu/~ccb/joshua/).
About speech generation from sign language, in the Computer Science department of the RWTH, Aachen University, P. Dreuw supervised by H. Ney is making a significant effort in recognizing continuous sign language with a new vision-based technology ([9]; [10]).
This paper describes the development of a Spanish Sign Language into Spoken Spanish translation system in a real domain: the Driver's License and Identity Document renewal. Specifically, the paper is focused on translation module between a sequence of written signs and written Spanish.

# 4. Database

In order to develop a translation system focused on the domain of renewal of ID and DL, a database has been generated. This database has been obtained with the collaboration of Local Government Offices where the mentioned services (renewal of ID and DL) are provided. During three weeks, the most frequent explanations (from government employees) and the most frequent questions (from the user) were taken down and more than 5,000 sentences were noted.
These 5,000 sentences were analysed because not all of them refer to ID or DL, so sentences were selected manually in order to develop a system in a specific domain. Finally, 1360 sentences were selected: 1,023 pronounced by government employees and 337 by users. These sentences were translated into LSE, both in text (sequence of glosses) and in video, and compiled in an excel file. This corpus was increased to 4,080 by incorporating different variants for Spanish sentences, maintaining the LSE translation.
The main features of the corpus are shown in Table 1.

| | **ID** | | **DL** | |
|---|---|---|---|---|
| **Government employee** | **Spanish** | **LSE** | **Spanish** | **LSE** |
| Sentence pairs | 1,425 | | 1,641 | |
| Different sentences | 1,236 | 389 | 1,413 | 199 |
| Running words | 8,490 | 6,282 | 17,113 | 12,741 |
| Vocabulary | 652 | 364 | 527 | 237 |
| **User** | **Spanish** | **LSE** | **Spanish** | **LSE** |
| Sentence pairs | 531 | | 483 | |
| Different sentences | 458 | 139 | 389 | 93 |
| Running words | 2,768 | 1,950 | 3,130 | 2,283 |
| Vocabulary | 422 | 165 | 294 | 133 |

Table 1: *Main statistics of the corpus.*

For the system development, two types of files were generated from the database: text files and sign files. Text files are composed of Spanish sentences of the parallel corpus and sign files contain their LSE translations (LSE sentences made up of glosses –capital words who represent signs).
These pairs of files were divided randomly into three sets: training (75%), development (12.5%) and test (12.5%), carrying out a round-robin evaluating process. The results presented in this paper are the average of this round robin, increasing the reliability of results this way.

# 5. Spanish generation from LSE

The spoken Spanish generation system converts a sign sequence (LSE sequence) into spoken Spanish. It is composed of three modules (Figure 1).

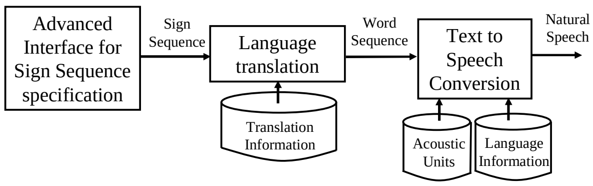


Figure 1: *Diagram of Spanish generation system.*

The first module is a visual interface for specifying the sign sequence. This interface includes several tools for sign specification: avatar for sign representation (to verify that sign corresponds to the gloss), prediction mechanisms, calendar and clock for date or time definitions, spelling, frequent questions, etc. With this visual interface the Deaf can build a sign sentence that will be translated into Spanish and spoken to a hearing person. The sign sequence is specified in glosses but signs can be searched by using specific sign characteristics in HamNoSys notation [12].
The second module converts a sign sequence into a word sequence with a statistical translation strategy.
The last module converts the word sequence into spoken Spanish by using a commercial Text to Speech converter. In this project, the Loquendo system has been used (http://www.loquendo.com/en/).
Visual interface is shown below in Figure 2.


Figure 2: *Visual interface for sign sequence specification.*

This paper describes the statistic translation system based on a phrase-based model.

### 5.1. Phrase-based translation

Phrase-based translation system uses a model of word sequences that is obtained from the alignment of parallel corpus (Figure 3). For the corpus alignment, *GIZA++* [13] program is used, and after, a set of phrases and their translation probabilities are obtained with *Phrase-Extract* and *Phrase-Score* programs.

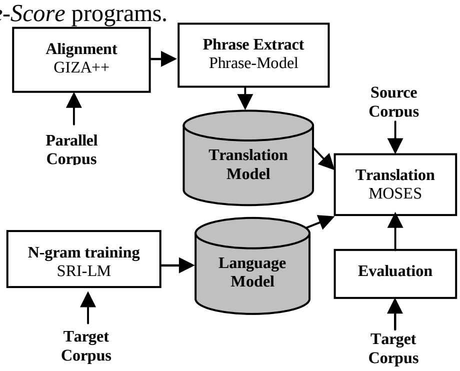


Figure 3: *Phrase-based translation module.*

Also, a 3-gram language model is incorporated, using the *SRI-LM* toolkit [14].

Translation model obtains *p(t|s)* (probability of *t (target language))* given *s (source language)*) with Bayes's theorem:

$$p(t|s) \approx p(s|t) \cdot p(t)$$

*p(s|t)* is the probability of *s* given *t* (translation model) and *p(t)* is the probability of seeing *t* (language model).

When translation and language models are generated, they have probability weights in translation that are not the best. Because of this, several translations are carried out with the development set in order to find probability weights that provide the best results. For translating, the *MOSES* decoder is used [15].

Finally, using the translation and language models and their probability weights, automatic translation is carried out with *MOSES* for evaluating the translation system.

### 5.2. Analysis of the alignment configuration

In order to generate the word alignment, *GIZA++* obtains the word alignment in both directions: source-target and target-source (LSE-Spanish and Spanish-LSE). Later, a final alignment is generated from a combination of previous alignments. Figure 4 shows different alignments between a pair of sentences in Spanish and LSE and their alignment points (each black box represents a word and a sign both aligned). The combination can be:

- **Source-Target (ST)**: Only the source-target (LSE-Spanish) alignment is considered. In this configuration, alignment is guided by signs: each sign in LSE is aligned with a Spanish word and it is possible that some word was unaligned.
- **Target-Source (TS)**: Target-source (Spanish-LSE) is the only considered alignment. In this configuration, alignment is guided by words: each Spanish word is aligned with a sign in LSE and it is possible that some sign was unaligned.
- **Union (U)**: In this case, alignment points of the union of both directions (source-target and target-source) are taken. This way, additional alignment points are obtained, having more examples for training the word translation model, but also, alignment quality is worse (more variability).
- **Intersection (I)**: In this case, alignment points of the intersection of both directions (source-target and target-source) are selected. This is the most strict configuration: less alignment points are obtained, but they are more reliable. This is not a good configuration if there is not a sufficient number of sentences for training.
- **Grow (G)**: In this configuration, alignment points of intersection are used to train the word translation model and also the adjoining points of union. This configuration is an intermediate solution between union and intersection, seeking a compromise between quality and quantity of alignment points.
- **Diagonal Grow (DG)**: In this configuration, alignment points of intersection are considered and also the adjoining points of union, but only adjoining points in diagonal.
- **Final Diagonal Grow (FDG)**: In this configuration, alignment points of intersection are taken and also the adjoining points of union, but only adjoining points in diagonal. And finally, if there is any word or sign unaligned, it is taken the corresponding union alignment point.

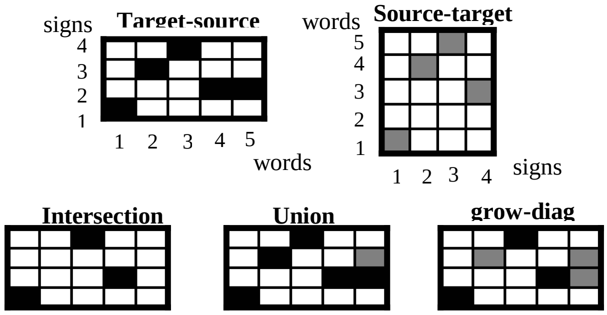


Figure 4: *Different alignment combinations.*

In order to analyse the effect of the alignment in final results, different alignment configurations were tested.

## 6. Evaluation

Table 2 and Figure 5 show the different results for each alignment configuration: mWER (multiple references Word Error Rate), BLEU (BiLingual Evaluation Understudy) and NIST. BLEU and NIST measures have been computed using the NIST tool (mteval.pl). Also, it is indicated the percentage of deletions (D), substitutions (S) and insertions (I) in translated sentences.

| | mWER (%) | D (%) | S (%) | I (%) | BLEU | NIST |
|---|---|---|---|---|---|---|
| **I** | 8.41 | 5.29 | 1.38 | 1.75 | 0.9252 | 11.7069 |
| **ST** | 6.52 | 4.28 | 1.09 | 1.14 | 0.9397 | 11.8033 |
| **DG** | 6.39 | 3.54 | 1.32 | 1.53 | 0.9430 | 11.8022 |
| **U** | 5.66 | 2.36 | 1.96 | 1.33 | 0.9459 | 11.7416 |
| **G** | 5.61 | 2.34 | 1.99 | 1.28 | 0.9459 | 11.7416 |
| **FDG** | 4.84 | 1.75 | 2.02 | 1.07 | 0.9520 | 11.7218 |
| **TS** | **3.90** | **1.68** | **1.34** | **0.89** | **0.9645** | **11.9020** |

Table 2: *Results of phrase-based translation system using different alignment configurations.*

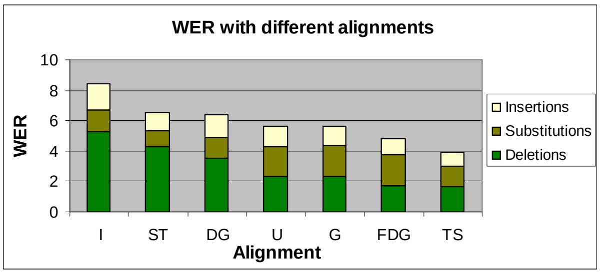


Figure 5: *Comparison of results of phrase-based translation system using different alignment configurations.*

As Table 2 and Figure 5 show, the best alignment is **target-source**: alignment is guided by words in this case. The main improvement is due to a less number of deletions and these deletions are important in translation because system translates from a language with less tokens per sentence (4.4 in LSE) to a language with more tokens per sentence (5.9 in Spanish). On the other hand, it can be observed that the worst result is given by intersection alignment, because important alignment points of target-source are deleted (looking at the Table 2, most mistakes are deletions). As additional points of target-source are added, results improve (deletions are reduced), and finally, with target-source the best result is obtained, giving a **3.90% mWER** and a **0.9645 BLEU**.

## 7. Conclusions

The paper has described the development of a spoken Spanish generation system from Spanish Sign Language. This system is focused on the application domain of renewal of Identity Document and Driver's license and it is composed of three modules: an interface for specifying the sign sequence, a statistical translation module for converting the sign sequence into a word sequence and a text to speech converter. The paper has been focused on this statistical translation module that uses a phrase-based translation model. Specifically, the paper has studied the alignment effect of the translation model in final results. The best alignment configuration is target-source where alignment is guided by words: less deletions in translated sentences are produced. Also, it can be observed that according as target-source alignment points are deleted, word error rate increases, because deletions in translated sentences increase and these deletions are important: system translates from LSE (with less tokens per sentence) to Spanish (with more tokens per sentence). Finally, it is obtained a 3.90% mWER and a 0.9645 BLEU.

## 8. Acknowledgements

The authors want to thank the eSIGN (Essential Sign Language Information on Government Networks) consortium for permitting the use of the eSIGN Editor and the 3D avatar in this research work. This work has been supported by Plan Avanza Exp Nº: PAV-070000-2007-567, ROBONAUTA (MEC ref: DPI2007-66846-c02-02) and SD-TEAM (MEC ref: TIN2008-06856-C05-03) projects.

## 9. References

[1] 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.
[2] Dreuw P., Neidle C., Athitsos V., Sclaroff S., and Ney H. 2008a. “Benchmark Databases for Video-Based Automatic Sign Language Recognition”. LREC, Marrakech, Morocco.
[3] 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.
[4] Efthimiou E., and Fotinea, E., 2008 “GSLC: Creation and Annotation of a Greek Sign Language Corpus for HCI” LREC 2008.
[5] Och J., Ney. H., 2002. “Discriminative Training and Maximum Entropy Models for Statistical Machine Translation”. Annual Meeting of the Ass. For Computational Linguistics (ACL), Philadelphia, PA, pp. 295-302. 2002.
[6] Mariño J.B., Banchs R., Crego J.M., Gispert A., Lambert P., Fonollosa J.A., Costa-Jussà M., 2006. "N-gram-based Machine Translation", Computational Linguistics, Association for Computacional Linguistics. Vol. 32, nº 4, pp. 527-549.
[7] Sumita E., Y. Akiba, T. Doi et al. 2003. “A Corpus-Centered Approach to Spoken Language Translation”. Conf. of the Europ. Chapter of the Ass. For Computational Linguistics (EACL), Budapest, Hungary. pp171-174. 2003.
[8] Casacuberta F., E. Vidal. 2004. “Machine Translation with Inferred Stochastic Finite-State Transducers”. Computational Linguistics, Vol. 30, No. 2, pp. 205-225, June 2004.
[9] Dreuw P., Ney H., Martinez G., Crasborn O., Piater J., Miguel Moya J., and Wheatley M., 2010 “The SignSpeak Project - Bridging the Gap Between Signers and Speakers”. In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010a.
[10] Dreuw P., Forster J., Gweth Y., Stein D., Ney H., Martinez G., Verges Llahi J., Crasborn O., Ormel E., Du W., Hoyoux T., Piater J., Moya Lazaro JM, and Wheatley M. 2010 “SignSpeak - Understanding, Recognition, and Translation of Sign Languages”. In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010b.
[11] Herrero, A. 2009. “Gramática didáctica de la Lengua de Signos Española (LSE)”. 2009.
[12] Prillwitz, S., R. Leven, H. Zienert, T. Hanke, J. Henning, et-al. 1989. “Hamburg Notation System for Sign Languages – An introductory Guide”. International Studies on Sign Language and the Communication of the Deaf, Vol. 5. Uni. of Hamburg.
[13] Och J., Ney H., 2000. "Improved Statistical Alignment Models". Proc. of the 38th Annual Meeting of the Association for Computational Linguistics, pp. 440-447, Hongkong, China.
[14] Stolcke A., 2002. “SRILM – An Extensible Language Modelling Toolkit”. Proc. Intl. Conf. on Spoken Language Processing, vol. 2, pp. 901-904, Denver.
[15] Koehn, Philipp. 2010. “Statistical Machine Translation”. Cambridge University Press.