Translating Bus Information into Sign Language for Deaf People

V. López-Ludeña , C. González-Morcillo , J.C. López , R. Barra-Chicote , R. Cordoba[1], R. San-Segundo

ABSTRACT

This paper describes the application of language translation technologies for generating bus information in Spanish Sign Language (LSE: Lengua de Signos Española). In this work, two main systems has been developed: the first for translating text messages from information panels, and the second for translating spoken Spanish into natural conversations at the information point of the bus company. Both systems are made up of a natural language translator (for converting a word sentence into a sequence of LSE signs), and a 3D avatar animation module (for playing back the signs). For the natural language translator, two technological approaches have been analyzed and integrated: an example-based strategy, and a statistical translator. When translating spoken utterances, it is also necessary to incorporate a speech recognizer for decoding the spoken utterance into a word sequence, prior to the language translation module. This paper includes a detailed description of the field evaluation carried out in this domain. This evaluation has been carried out at the customer information office in Madrid involving both real bus company employees and deaf people. The evaluation includes objective measurements from the system and information from questionnaires. In the field evaluation, the whole translation presents an SER (Sign Error Rate) of less than 10% and a BLEU greater than 90%.

CONTENTS

1. Introduction

According to wfdeaf.org (2013), there are more than 70 million deaf people in the world. This disability has serious implications for education and social inclusion. In Spain, there are 1,064,000 deaf people according to the INE (Spanish Institute of Statistics). 47% of the deaf population do not have basic studies or are illiterate, and only between 1% and 3% have finished their university studies (as opposed to 21% of Spanish hearing people). Deaf people (especially those that became deaf before language acquisition) have serious problems when expressing themselves or understanding written texts. They have problems with verb tenses, concordances of gender and number, etc., and they have difficulties when creating a mental image of abstract concepts. These deficiencies have become apparent because of the lack of feedback in speak-listen procedures.

However, the Deaf[1] use a sign language (their mother tongue) for communicating. Sign languages are fully-fledged languages that have a grammar and lexicon just like any spoken language, contrary to what most people think. Traditionally, deafness has been associated with people with learning problems but this is not the case. The use of sign languages defines the Deaf as a linguistic minority, with learning skills, cultural and group rights similar to other minority language communities. 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. LSE is a natural language with the same linguistic levels as other languages such as Spanish. Thanks to associations such as the Fundación CNSE, LSE is becoming the natural language for the Deaf to communicate.

This paper describes the efforts made to translate transport information into LSE, specifically bus information. The main target is to translate this information automatically (without human intervention).

This paper is organised as follows. Section 2 presents the state of the art. Section 3 describes the main language translation technologies considered in this work. The sign representation using an animated avatar is described in section 4. Section 5 describes the system for translating panel information, and section 6 the system for translating face-to-face conversations at the customer service office, including a field evaluation. Finally, section 7 summaries the main conclusions of this work.

2. State of the art

In the last 20 years, the European Commission and the USA Government have invested many resources into research into language translation. In Europe, there has been a large sequence of research projects: C-Star, ATR. Vermobil, Eutrans, LC-Star, PF-Star and, finally, TC-STAR, EuroMatrix, EuroMatrixPlus, FAUST, etc. Some of them focus on text translation and others on spoken language. The FAUST project focuses on computer-aided translation. In the USA, DARPA (Defense Advanced Research Projects Agency) is supporting the GALE program (http://www.darpa.mil/ipto/programs/gale/gale.asp). The goal of the DARPA GALE program has been to develop and apply computer software technologies to absorb, analyze and interpret huge volumes of speech and text in multiple languages. This program has also been promoted by the machine translation evaluation organised by the US Government, NIST (http://www.itl.nist.gov/iad/mig/tests/mt/).

The best performing translation systems are based on various types of statistical approaches (Och and Ney, 2002; Mariño et al, 2006), including example-based methods (Sumita et al, 2003), finite-state transducers (Casacuberta and Vidal, 2004) and other data-driven approaches. The progress made over the last 10 years is due to several factors such as efficient algorithms for training (Och and Ney, 2003), context dependent models (Zens et al, 2002), efficient algorithms for generation (Koehn, 2003), more powerful computers and bigger parallel corpora, and automatic error measurements (Papineli et al, 2002; Banerjee and Lavie, 2005; Agarwal and Lavie, 2008).

Another important effort in machine translation has been the organization of several Workshops on Statistical Machine Translation (SMT). On the webpage http://www.statmt.org/, it is possible to obtain all the information on these events. As a result of these workshops, there is a free machine translation system called Moses available from this web page (http://www.statmt.org/moses/). Moses is a phrase-based statistical machine translation system that allows you to build machine translation system models for any language pair, using a collection of translated texts (parallel corpus).

---

[1] It is necessary differentiate between "deaf" and "Deaf": the former refers to non-hearing people, and the latter refers to non-hearing people who use a sign language to communicate between themselves (their mother tongue), making them part of the "Deaf community".

In recent years, several groups have shown interest in spoken language translation into sign languages, developing several prototypes: example-based (Morrisey and Way, 2005), rule-based (San-Segundo et al 2008), grammar-based (Marshall and Sáfár, 2005), full sentence (Cox et al, 2002) or statistical (Bungeroth and Ney, 2004; SiSi system http://www-03.ibm.com/press/us/en/pressrelease/22316.wss; Morrissey et al, 2007) approaches. For LSE, it is important to highlight the authors' experience in developing speech into LSE translation systems in several domains (San Segundo et al., 2008; San Segundo et al., 2011; López-Ludeña et al, 2011; López-Ludeña et al, 2013a). This kind of system can complement a Sign Language into Speech translation system, allowing a two-direction interaction (Cemil et al, 2011; Ibarguren et al, 2010).

As regards 3D avatars for representing signs, the VISICAST and eSIGN European Project (Essential Sign Language Information on Government Networks) (http://www.sign-lang.uni-hamburg.de/esign/) (Zwiterslood et al, 2004) has been one of the most significant research efforts into developing tools for the automatic generation of sign language contents. One of the partners in the VISICAST and eSIGN projects is the research group into Virtual Humans at the University of East Anglia (http://www.uea.ac.uk/cmp/research/graphicsvisionspeech/vh). This group has been involved in several projects as regards the generation of sign language using virtual humans: TESSA, SignTel, Visicast, eSIGN, SiSi, LinguaSign, etc.

This paper describes the effort in adapting translation technology for generating LSE content into the bus information domain. This technology has been used for translating both panel information and spoken Spanish into LSE in real interactions between a deaf person and a hearing person without an interpreter: deaf customers and bus company employees that provide bus information. The method and the system used in this research work, has been developed during several years in previous research projects (San Segundo et al., 2008; San Segundo et al., 2011; López-Ludeña et al, 2011; López-Ludeña et al, 2013a).

3. Language translation technology

In this work, several translation strategies (López-Ludeña et al, 2013a) have been adapted and evaluated in the bus information domain: example-based and statistical translation. The final translation module integrates all these technologies.

In order to use automatic translating technologies, it is essential to represent the Sign Language in a written form. In order to write down LSE, each sign of an LSE sentence is represented by a gloss, so a gloss sequence represents a sequence of signs. Glosses are words in capital letters with a similar meaning to the sign meaning. An example of glosses representing the sentence "¿a qué hora se abre? (what time do you open?)" would be "ABRIR HORA? (OPEN HOUR?)". There can be several signs represented by a gloss with '+', for example: "SABADO+DOMINGO (SATURDAY+SUNDAY)" to represent "fin de semana (weekend)". Also, there can be several words in Spanish that form only one gloss in LSE, this fact is marked with '-'. For example, "CAFE-CON-LECHE" for representing "café con leche (coffee with milk)". For more details about LSE and written LSE can be seen at (López-Ludeña et al, 2011).

3.1. Example-based strategy

An example-based translation system uses a parallel corpus: set of sentences in the source language (from which one is translating) and its corresponding translations into the target language, and translates other similar source-language sentences. In order to determine whether one example (in the corpora) is similar enough to the text to be translated, the system computes a heuristic distance between both sentences. If the distance is less than a threshold, the translation output will be the same as the example translation. But if the distance is greater, the system cannot generate any output and it is necessary to consider other translation strategies.

In this case, the heuristic distance considered is the well-known Levenshtein distance (LD) (Levenshtein, 1966) divided by the number of words in the sentence to be translated (this distance is represented as a percentage). The Levenshtein Distance is a measurement of the similarity between two strings (or character sequences): source sequence (s) and target sequence (t). The distance is the number of deletions, insertions, or substitutions required to transform s into t. Because of this, it is also called the edit distance. Originally, this distance was used to measure the similarity between two strings (character sequences). But it was already used for defining a distance between word sequences (as has been used in this paper). The LD is computed using a dynamic programming algorithm that considers the following costs: 0 for identical words, 1 for insertions, 1 for deletions and 1 for substitutions.

In order to develop an example-based translation system, it is necessary a large amounts of pre-translated text to make a reasonable translator. But it is possible to generalize the examples in order to make them more effective:

more than one string can match any given part of the example. Considering the following translation example for Spanish into LSE:

Spanish: "Veinte euros con diez céntimos" (Twenty Euros, ten cents)

LSE: "VEINTE COMA DIEZ EURO"

Now, if it is known that "veinte" and "diez" are numbers, it is possible to save this example in the corpus as

Spanish: "$NUMBER euros con $NUMBER céntimos"

LSE: "$NUMBER COMA $NUMBER EURO"

where $NUMBER is a word class including all numbers. Notice how it is possible to match many other strings that have this pattern. They are not restricted to these numbers. When indexing the example corpora, and before matching a new input against the database, the system tags the input by searching for words and phrases included in the class lists, and replacing each occurrence with the appropriate token. There is a file which simply lists all the members of a class in a group, along with the corresponding translation for each token. For the system implemented, 4 classes were used: $NUMBER, $PROPER_NAME, $MONTH and $WEEK_DAY.

3.2. Statistical translation

For statistical translation, two methods have been evaluated: a Phrase-based Translator and a Stochastic Finite State Transducer (SFST). The phrase-based translation system is based on the software released from NAACL Workshops on Statistical Machine Translation (http://www.statmt.org). The translation process uses a translation model based on phrases and a target language model.
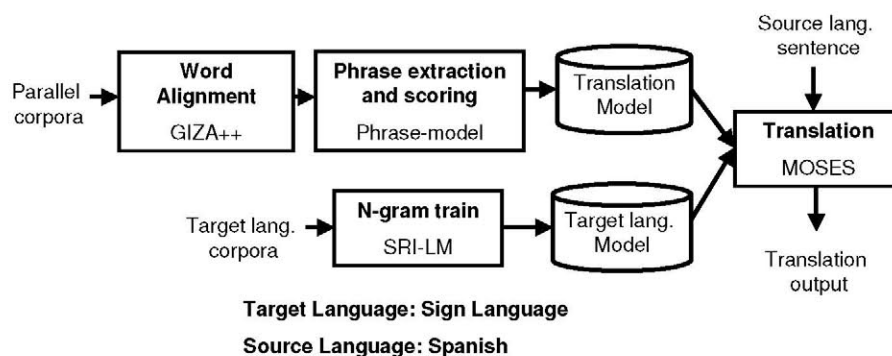

Figure 1. Diagram of the phrase-based translation module

The phrase model has been trained using the following steps (Figure 1):

- Word alignment computation. In this step, the GIZA++ software (Och and Ney, 2000) has been used to calculate the alignments between words and signs. In order to establish word alignments, GIZA++ combines the alignments in both directions: words-signs and signs-words. GIZA++ also generates a lexical translation model including the translation probability between every word and every sign.

- Phrase extraction (Koehn et al 2003). All phrase pairs that are consistent with the word alignment are collected. For a phrase alignment to be consistent with the word alignment, all alignment points for rows and columns that are touched by the box have to be in the box, not outside (Figure 2). The maximum size of a phrase has been fixed at 7.
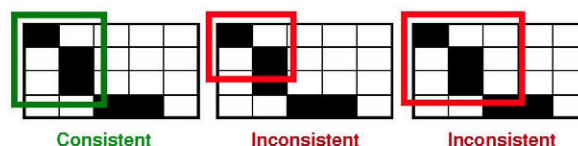

Figure 2. Examples of phrase extraction.

- 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 (http://www.statmt.org/moses/) 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 needed by Moses, the SRI language modelling toolkit has been used (Stolcke, 2002).
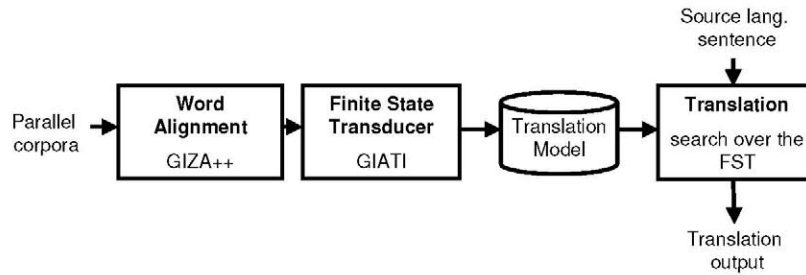

Figure 3. Diagram of the FST-based translation module

The translation based on SFST is carried out as detailed in Figure 3. The translation model consists of an SFST made up of aggregations: subsequences of aligned source and target words. The SFST is inferred from the word alignment (obtained with GIZA++) using the GIATI (Grammatical Inference and Alignments for Transducer Inference) algorithm (Casacuberta and Vidal, 2004). The SFST probabilities are also trained from aligned corpora. The software used in this paper has been downloaded from http://prhlt.iti.es/content.php?page=software.php.

Both statistical translation strategies incorporate a new pre-processing module (López-Ludeña et al, 2011) that permits its performance to be increased.

3.3.   Integrating translation strategies

The natural language translation module implemented combines the two translation strategies described in previous sections. This combination is detailed in Figure 4.

The translation module has a hierarchical structure divided into two main steps. In the first step, an example-based strategy is used to translate the word sequence. If the distance with the closest example is less than a certain threshold (Distance Threshold), the translation output is the same as the example. But if the distance is greater, a background module translates the word sequence. The Distance Threshold (DT) ranges between 20% and 30%. In the field evaluation, the DT was fixed at 30% (one difference is permitted in a 4-word sentence).
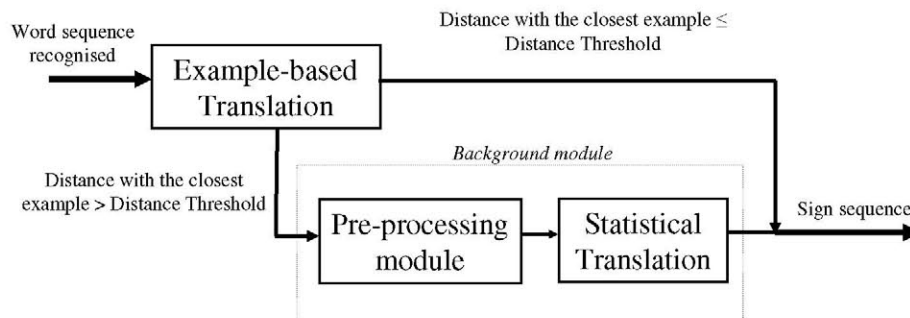

Figure 4. Diagram of natural language translation module combining two different translation strategies

For the background module, both alternatives of statistical translation were incorporated (phrase-based and SFST-based strategies), although only the phrase-based one was used for the field evaluation because of its better performance (as it will be shown latter).

3.4.   Laboratory evaluation

In order to develop and evaluate the translation technology, it is necessary to generate a parallel corpus with Spanish sentences and their translation into LSE. This generation process consists of two main steps:

* First, it is necessary to collect the Spanish sentences from dialogues between customers and bus company employees. This collection has been obtained with the collaboration of the bus company in Madrid. Over several days, the most frequent explanations (from the bus company employee) and the most frequent questions (from the customer) were compiled. During this period, more than 1,500 sentences were taken

down and analysed. Not all the sentences refer to information services, so the sentences had to be selected manually. This was possible because every sentence was tagged with the information on the service being provided when it was collected. Finally, 500 sentences were compiled: 289 pronounced by bus company employees and 211 by customers. This corpus was increased to 1,938 by incorporating different variants for Spanish sentences (maintaining the meaning and the LSE translation).

- These sentences were translated into LSE, both in text (sequence of glosses) and in video, and compiled in an excel file. The excel file contains eight different information fields: "INDEX" (sentence index), "DOMAIN" (bus information in this case), "SCENARIO" (scenario: where the sentence was collected), "SERVICE" (service provided when the sentence was collected), if the sentence was pronounced by the bus company employee to the customer (AGENT), sentence in Spanish (SPANISH), sentence in LSE (sequence of glosses), and link to the video file with LSE representation. The main features of the corpus are summarised in Table 1. These features are divided whether the sentence was spoken by the bus company employee or the customer.

Bus company employee	Spanish	LSE
Sentence pairs	1110	
Different sentences	906	292
Running words	7289	4158
Vocabulary	698	348
Customer	Spanish	LSE
Sentence pairs	828	
Different sentences	627	219
Running words	4139	2748
Vocabulary	527	285

Table 1. Main statistics of the parallel corpus

In order to evaluate the different translation approaches, a Cross-Validation process was carried out. The corpus (including only those sentences pronounced by bus company employees: Table 1) was divided randomly into three disjoint sets: training (75% of the sentences), development (12.5% of the sentences) and test (12.5% of the sentences). This way, these translation technologies were trained, tuned and tested using disjoin sets. The translation results were computed over the test set. This experiment was repeated 8 times changing the set division based on a round-robin strategy. Table 2 presents the average translation results over these 8 experiments.

Table 2 summarizes the results for example-based and statistical approaches considering several performance metrics: SER (Sign Error Rate) is the percentage of wrong signs in the translation output compared to the reference in the same order. PER (Position Independent SER) is the percentage of wrong signs in the translation output compared to the reference without considering the order. BLEU (BiLingual Evaluation Understudy; (Papineni, 2002)) is an algorithm for evaluating the quality of an automatic translation. The main task is to compare n-grams (sequences of n signs) of the translation output with the n-grams of the reference translation and count the number of matches. These matches are position independent. The more the matches, the better the candidate translation is. BLEU was one of the first metrics to achieve a high correlation with human judgements of quality. BLEU's output is always a number between 0 and 1. This value indicates how similar the candidate and reference sentences are; values closer to 1 represent more similar sentences. It is important to highlight that SER and PER are error metrics (a lower value means a better result) while BLEU is an accuracy metric (a higher value means a better result).

	SER (%)	±Δ	PER (%)	BLEU (%)
Statistical approach: phrase-based	31.89	1.08	29.67	66.72
Statistical approach: SFST-based	34.30	1.10	32.29	61.58
Example-based approach	33.31	1.10	31.45	65.12
Example-based approach (considering a heuristic distance < 30%)	5.34	0.53	4.29	93.12
Combining translation strategies	7.70	0.62	6.34	91.56

Table 2. Result summary for example-based, rule-based and statistical approaches

For every SER result, the confidence interval (at 95%) is also presented. This interval is calculated using the following formula:

$$\pm \Delta = 1{,}96\sqrt{\dfrac{SER\ (100 - SER)}{n}}$$

Equation 1. Confidence Interval at 95%

n is the number of signs used in testing. An improvement between two systems is statistically significant when there is no overlap between the confidence intervals of both systems. As shown in Table 2, all improvements between different approaches are higher than the confidence intervals.

As shown in Table 2, for this corpus, the phrase-based and example-based methods are better than the SFST-based method.

Table 2 also presents the translation results for the example-based approach for those sentences that have a heuristic distance (with the closest example) of less than 30% (the rest of the sentences were not translated). In this case, the results increase significantly: SER improvement is greater than the confidence intervals (at 95%). Finally, Table 2 presents the results for the combination of several translation strategies: example-based (considering a heuristic distance < 30%) and phrase-based approaches. As is shown, with the hierarchical system it is possible to obtain better results by translating all the test sentences: SER < 10%.

The hierarchical module has been used in the field evaluation (section 6). For the field evaluation, the translation module has been trained with all the information in the database described in Table 1.

4.   Sign Language Representation

The animation module uses a declarative abstraction module used by all of the internal components. This module uses a description based on XML, where each key pose configuration is stored defining its position, rotation, length and hierarchical structure. We have used an approximation of the standard defined by H-Anim (Humanoid Working Group ISO/IEC FCD 19774:200x). In terms of the bones hierarchy, each animation chain is made up of several « joint » objects that define transformations from the root of the hierarchy.

Several general purpose avatars such as Greta (Niewiadmonski et al., 2009) or SmartBody (Thiebaux et al., 2008) have lacked a significant number of essential features for sign language synthesis. Hand configuration is an extremely important feature; the meaning of a sign is strongly related to the finger position and rotation. In our avatar each phalanx can be positioned and rotated using realistic human limitations. This is the most time-consuming phase in the generation of a new sign and, as detailed in the following section; a new approach to increase the adaptability has been created. For each sign it is necessary to model non-manual features (torso movements, facial expressions and gaze). The skeleton defined in the representation module is made up of 103 bones, out of which 19 are inverse kinematics handlers (they have an influence on a set of bones). The use of inverse kinematics and spherical quaternion interpolation (Watt and Watt, 1992) eases the work of the animators in capturing the key poses of signs from deaf experts. The geometry of the avatar is defined using Catmull-Clark adaptive subdivision surfaces. To ease the portability for real time rendering, each vertex has the same weight (each vertex has the same influence on the final deformation of the mesh).

There are three main concepts related to inverse kinematics methods: the description of the joints, the rotation angle and the degrees of freedom. The joints' own physical features that determine the final movement, the rotation angle describes the allowed rotation for the point of union and the degrees of freedom involve the directions in which a joint moves. In most kinematics configurations it is essential to define rotation constraints to avoid forbidden configurations and simulate only physically-correct positions. There are two ways of dealing with IK: analytic or iterative methods. The analytic methods require a previous analysis of the animation hierarchy and, in the case of complex configurations (such as virtual avatars), the resulting equations can be quite complex and computationally intensive. To overcome this problem, this module uses the Cyclic Coordinate Descent CCD algorithm (Lever, 2002). CCD is an iterative method to compute IK that minimizes the error of the kinematic configuration for each joint. The algorithm starts computing the rotation of the first element of the chain and iterates the elements, adjusting the configuration of each joint until the position of the effector is close to the desired position, or a specific number of iterations is reached.

Facial expression is used to indicate the sentence mode (assertion or question) and eyebrows are related to the information structure. In this way, this non-manual animation is used to highlight adjectival or adverbial information. The movements of the mouth are also highly important in focusing the visual attention to make

comprehension easier. As pointed out by Pfau (Pfau and Quer, 2010), non-manuals require more attention from the point of view of the automatic sign language synthesis.

The composition of the final animation of the character is based on Non-Linear Animation techniques (NLA) (Lever, 2002). NLA techniques are used in film production to merge individual actions into complex animations. Each small piece of animation (action) is specialized in one thing. These actions can easily be reused in different domains. Thanks to the use of this approach, each action defines an animation layer (such as body, hand or face animation). Each sign is defined by means of several actions (or animation channels, e.g. Facial, Hands or Modifiers). The final movement of the sign is obtained by fusing the described animation layers. For instance, there are three basic actions defined in Figure 5 to create a « question about a big cat ». Basic SLERP interpolation (Watt and Watt, 1992) is also used to concatenate signs smoothly in an utterance.
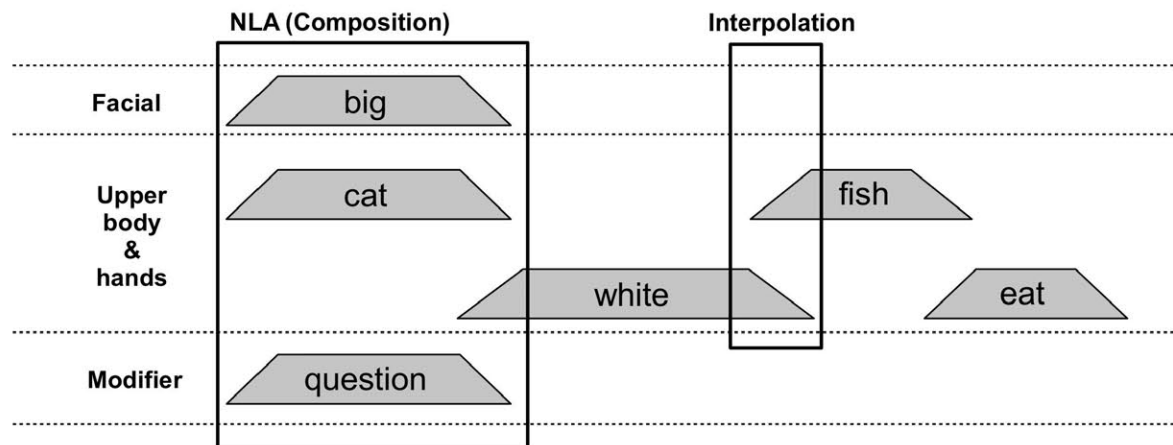

Figure 5. Composition of animation channels (NLA) and interpolation between signs.

The realistic result of the movements is probably the most important elements to consider in the representation of sign language. The results obtained in this work improve the results obtained in similar systems thanks to the use of the realistic rendering approach and the composition of individual actions. The key frame animation approach produces more accurate, comprehensible and lifelike results than motion capture-based techniques (Adamo-Villani, 2008).

Another advantage of the representation module is the adaptation to different kinds of devices (computers, mobile phones, etc). The rendering phase is often considered as a bottleneck in photorealistic projects in which one image may need hours of rendering in a modern workstation. The rendering system used in this work can be easily used through distributed rendering approaches (Gonzalez-Morcillo et al., 2010).

Social responses to virtual humans have been studied using both objective and subjective methods in different contexts. The behavioural realism of their movements has a strong effect on the quality of communication in general, and in the subjective impression of understanding in sign language in particular (Kipp et al., 2011). Depending on the application domain (the gender, age and cultural awareness of the final user), the representation of the avatar must be changed. To avoid the rejection of the final user, this form of adaptability is needed in any real-world scenario. In this work, the representation of the internal IK skeleton is shared between virtual characters using an XML specification. This file also specifies the relative size of the bones and the constraints required to generate realistic movements. Figure 6 shows an example of the reuse of the same pose.

Another important factor to increase the adaptability is the generation of the specific vocabulary in each application domain. Thanks to the use of an internal skeleton shared between avatars, the definition of each sign need only be made once. In previous developments of this representation module (Herrera et al., 2009), the movement description of each sign was done by trained experts in computer animation and sign language. Using a real video of a native signer, the expert detected the relevant changes in the direction of the joints adding key frames using the appropriate rotation value.

Figure 6. Two results using different avatars (Niva and Perico) with realistic rendering and the same pose.

4.1. Sign Editor description

One of the main problems related to the creation of the signs is the time required for modelling. In spite of the development of new techniques to facilitate the animation of virtual characters (such as inverse kinematics controls and key poses), the user may spend between 15 and 30 minutes setting a new sign. It is important to recall that each sign must be made only once and thanks to the design of the representation module, this description of the movement can be reused in different 3D avatars. Because of the huge amount of time required, this phase may be considered the main bottleneck in the project.

Figure 7. Sign editor based on the use of predefined static poses for hand shape and orientations.

A sign editor module (Figure 7) has been developed to ease the construction of the sign dictionary. In this application, the user chooses basic configurations of shape and orientations of the both hands (active and passive). The expert chooses the frame and with one interaction picks the closest configuration of the hand. This configuration can be refined later using the aforementioned inversed kinematics facilities. These configurations of the shape and orientation are defined as static poses which contain only the essential parameters that describe the action. This information is stored in XML files. Figure 8 presents the interface of the orientation panel and the description of the fifth pose.

In the current system, 86 hand shapes (23 basic shapes and 63 derived from the basic configurations) were defined. 53 configurations for orientation were also constructed. Thanks to the use of this sign editor, the time required to specify a new sign decreased by 90% with similar quality results. Some examples can be downloaded from http://www.esi.uclm.es/www/cglez/ConSignos/listadoSignos/.

```xml
<?xml version="1.0" encoding="utf-8"?>
<con_signos>
  <pose armature="Armature"/>
  <pose action="/orientation/right/5"/>
  <pose pose_name="/orientation/right/5">
    <bone name="hand.r">
      <quaternion id="quat_0"> 0.766044318676 </quaternion>
      <quaternion id="quat_1"> 1.26913831693e-16 </quaternion>
      <quaternion id="quat_2"> -0.64278781414 </quaternion>
      <quaternion id="quat_3"> -4.2781725033e-17 </quaternion>
    </bone>
    <bone name="ikhand.r">
      <quaternion id="quat_0"> 1.0 </quaternion>
      <quaternion id="quat_1"> 0.0 </quaternion>
      <quaternion id="quat_2"> 0.0 </quaternion>
      <quaternion id="quat_3"> 0.0 </quaternion>
      <location id="loc_0"> -1.55628681183 </location>
      <location id="loc_1"> -3.97473931313 </location>
      <location id="loc_2"> -0.200374856591 </location>
    </bone>
  </pose>
</con_signos>
```
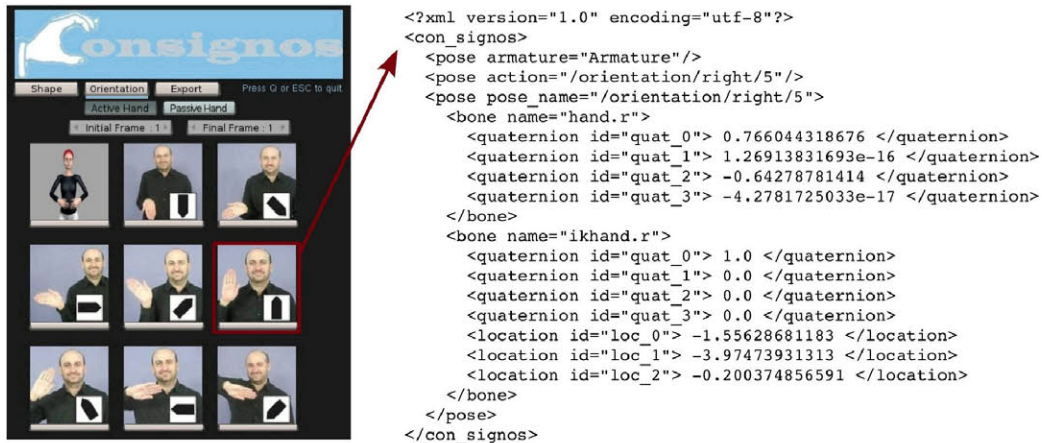
Figure 8. Static pose definition in XML of a right hand orientation setup.

4.2. Efficiency of the Sign Language representation module

In order to evaluate the performance of the sign representation module independently of the translation process (without translation errors), several tests were performed considering correct sentences in LSE. Several sentences from the corpus described in section 3.4 (Table 1) were randomly selected and presented to ten deaf customers (five females and five males). In the experiments, eight short sentences and eight long sentences were presented to each user (80 short sentences and 80 long sentences in total). They were asked to identify the sentences considering the specific domain. Table 4 summaries the recognition accuracy for short (less than four signs) and long messages (more than three signs) based on the number of attempts: number of times it was necessary to represent the sentence in LSE for being recognised. The recognition accuracy includes all the experiments with the ten deaf customers. As it is shown, all the sentences were recognised correctly after 3 times. Also higher recognition rate is obtained for shorter sentences.

Human recognition rate depending on the number of times the LSE sentence was represented			
	1st	2nd	3rd
Short messages (< 4 signs): 80 sentences	91.25%	96.25%	100.0%
Long messages (>= 4 signs): 80 sentences	87.50%	95.00%	100.0%

Table 3: Recognition accuracy based on the number of times the sentences were represented. Percentages represent accumulative recognition rates. In these experiments, there were not translation errors.

Regarding the processing time, it is important to comment that the Sign Language Representation module works in real time: the time for rendering is lower that the time for presenting the frames.

5. Translating panel information

Figure 9 shows several examples of bus information provided though panels situated at bus stops. Generally, these panels provide information on how long the customer must wait for the next bus. The panel shows the line (there may be several lines at the same stop), the destination and the number of minutes until the next bus. In this situation, Deaf people do not need any translation because the information is easy to understand in written Spanish. The problem appears when these panels are used to provide additional information like information about a strike, an accident on one line or other possible problems on other lines in the network. In this case, it is very useful to translate these long messages into LSE.

Figure 9. Examples of bus information provided through different panels

The translation system is made up of two main modules: the language translator described in section 3.3 and the sign representation module (section 4).
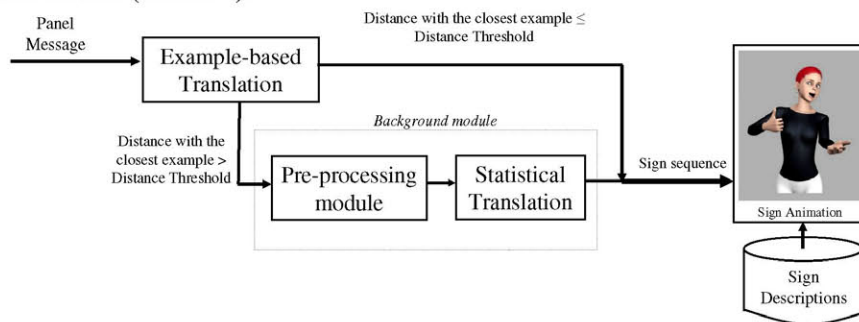

Figure 10. Diagram of the system for translating panel messages

This system was evaluated by ten deaf customers (five females and five males) using new messages, not considered in the laboratory experiments (section 3.4). The deaf customers observed 10 short messages and 10 long messages (100 short messages and 100 long messages in total) and they were asked to identify them considering the specific domain. Table 4 summaries the recognition accuracy for short (less than four signs) and long messages (more than three signs) based on the number of times it was necessary to represent the sentence in LSE for being recognised.

Human recognition rate depending on the number of times the LSE sentence was represented			
	1st	2nd	3rd
Short messages (< 4 signs): 100 messages	85.0%	96.0%	100.0%
Long messages (>= 4 signs): 100 messages	78.0%	95.0%	100.0%

Table 4: Recognition accuracy based on the number of times the sentences were represented. Percentages represent accumulative recognition rates.

As it is shown, although all the sentences were recognised correctly after 3 times, the recognition rates obtained the first time were lower than those presented in Table 3. These results are because small translation errors make the recognition more difficult. Spite of these errors, the user can understand the whole sentence (after several times) using contextual information. When considering short messages, the recognition rate is higher with fewer attempts. Finally, it is important to comment that, in order to represent sign language, the information panel should be replaced with a high resolution screen. The bus company in Madrid is currently working on this replacement.

6. Translating online conversations at the face-to-face customer service

The second translation system focuses on translating bus company employee utterances when providing information to customers. In this scenario, a two-directional translation system was developed to enable a dynamic conversation between a bus company employee and a deaf customer. Figure 11 shows an example of interaction at the customer service office.

Figure 11. Example of interaction at the customer service office.

6.1. System Overview

For dealing with face-to-face conversations, the translation system must contain two main modules: a speech into LSE translation module and a speech generation from LSE.
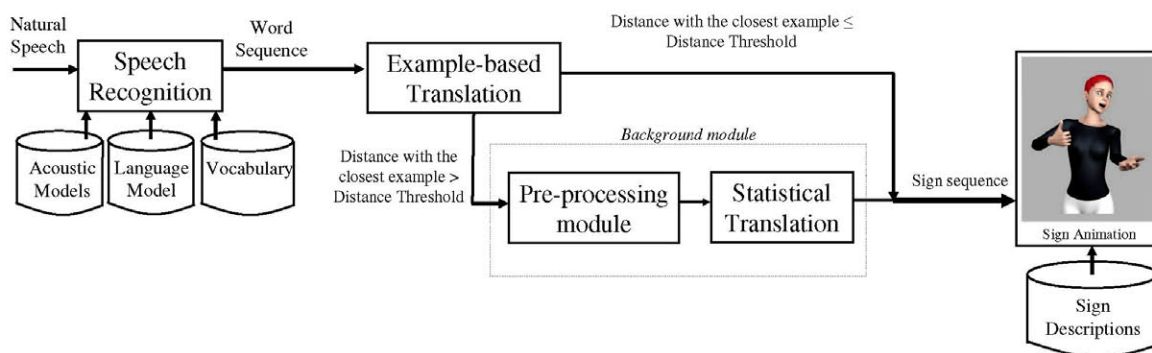
Figure 12. Module Diagram of the speech into LSE translation system

Figure 12 shows the diagram of the speech into LSE translation system. This system is used to translate spoken explanations from the bus company employee. This system is made up of three main modules:

- The automatic speech recognizer (ASR) converts natural speech into a sequence of words (text). It uses a vocabulary, a language model and acoustic models for every allophone.
- The natural language translation module converts a word sequence into a sign sequence. This module combines two different strategies (section 3.3). The first consists of an example-based strategy: the translation process is carried out based on the similarity between the sentence to be translated and the examples of a parallel corpus (examples and their corresponding translations). The second is based on a statistical translation approach where parallel corpora are used for training language and translation models.
- At the final step, the sign animation is made by using a highly accurate representation of the movements (hands, arms and facial expressions) in a Sign list database and a Non-Linear Animation composition module, both needed to generate clear output (section 4). This representation is independent of the virtual character and the final representation phase. In this way, the virtual character can be easily changed and the results can be adapted for use in different devices.

In order to convert deaf customer questions into spoken Spanish, the LSESpeak system was used (López-Ludeña et al, 2013b). LSESpeak is made up of two main tools (Figure 13). The first tool is a new version of an LSE into Spanish translation system (San-Segundo et al, 2010), and the second is an SMS to Spanish translation system, because Spanish deaf people have become familiar with SMS language. Both tools are made up of three main modules. The first module is an advanced interface in which it is possible to specify an LSE sequence or an SMS message. Secondly, there are two language translators for converting LSE or SMS (respectively) into written Spanish. Finally, the third module is an emotional text to speech converter in which the user can choose the voice gender (female or male), the emotion type (happy, sad, angry, surprise, and fear) and the Emotional Strength (ES) (on a 0-100% scale).
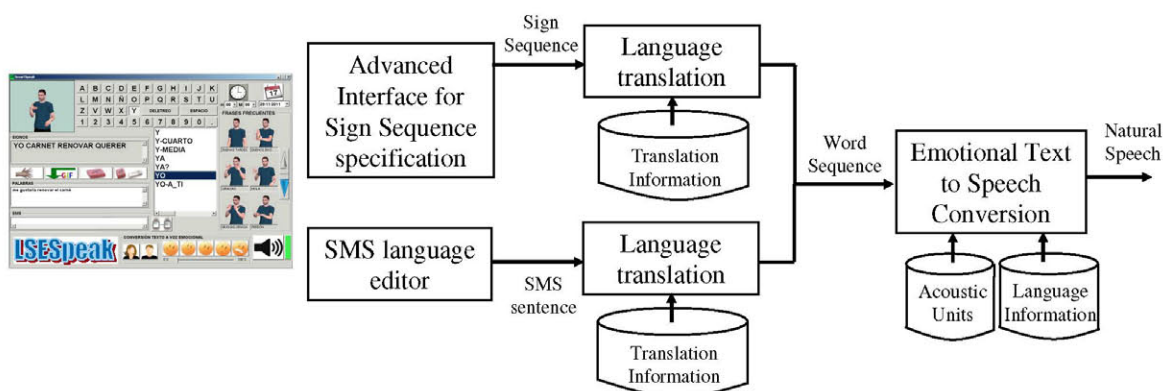

Figure 13. Module Diagram of LSESpeak

The LSE into written Spanish translation module have the same structure described in section 3.3. This module has been trained with the parallel corpus reported in section 3.4 (Table 1). In this case, only those sentences pronounced by customers have been considered.

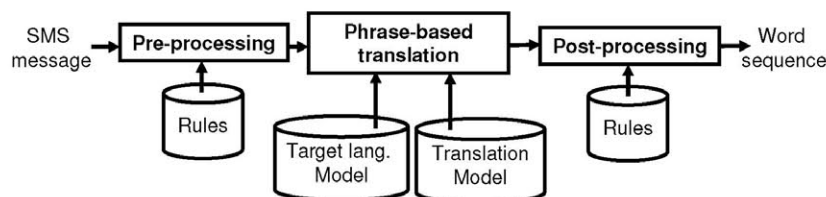The SMS into written Spanish translation module is represented in Figure 14.

Figure 14: Diagram of SMS to Spanish translation system.

First of all, there is a pre-processing module that prepares the SMS sentence before sending it to the automatic translator. The pre-processing module checks if there is any question or exclamation mark and, if so, to remove it from the sentence and mark that fact (with the activation of a flag) in order to take it into account in the post-processing. Secondly, the pre-processing checks if there is any special character like '+' or '#' next to any term and, if so, the system introduces a space between the character and the term. This action is necessary because, generally, these two isolated characters are translated by the Spanish words "más" (more) and "número" (number), respectively. For example, "q+ kiers?" would be translated into "¿Qué más quieres? (What else do you want?)".

The second module is a statistical translation system that consists of a phrase-based translator (Moses http://www.statmt.org/moses/, the same as that explained in the section 3.2). This statistical translation module has been trained with a dictionary of terms extracted from www.diccionariosms.com. This dictionary has been generated by Internet users. This dictionary contains more than 11,000 terms and expressions in SMS language (although this number increases every day) with their Spanish translations and a popularity rate based on the number of users who have registered the term-translation pair.

The third one is a post-processing module. The sentence translated by the phrase-based translator may contain some SMS terms that have not been correctly translated by Moses. In order to detect these terms the post-processing module check if every term, in the translated sentence, is pronounceable or not (according to the sequence of consonants and vowels in Spanish). If it is not pronounceable, the term is replaced by the most similar Spanish word (considering the Levenshtein distance). Finally, when the translation of the sentence is complete, it is necessary to check whether the sentence is interrogative or exclamatory (indicated by a flag) to add or not question or exclamation marks at the beginning and at the end of the sentence. More details can be seen in the LSESpeak paper (López-Ludeña et al, 2013b).

6.2. Field Evaluation

The information point is situated in the street, as shown in Figure 11. In order to avoid disturbing the normal working of this information point, the evaluation was carried out in a meeting room inside the office. Every evaluation session started with a one-hour talk about the project and the evaluation process given to bus company employees and deaf customers involved in the evaluation (Figure 15). It is important to remark that for this field evaluation, new dialogues were considered, different from those presented in the laboratory evaluation (section 3.4) or in the panel information evaluation (section 5).

Figure 15. Different photos at the customer service office during the evaluation

The system was evaluated by ten deaf customers (five female and five male) who interacted with three bus company employees at the information point during two different evaluation sessions. Every evaluation session lasted more than 5 hours. First of all, the deaf customers looked at several signs (10 signs per user) and were asked to identify them considering the specific domain. After that, they were asked to interact with the bus company employees using the translation systems in five different scenarios: in four of them, the customers

asked for information about buses going to a specific place (hospital, official building or tourist monument). In the other scenario, the customers asked about a lost object.

After the interactions, the deaf customers were asked specific questions about the information provided by the bus company employees. Traditionally, subjective measurements have been obtained by means of questionnaires filled in by the users in which several aspects related to the system performance are asked to the user in a general sense (for example, Is the translation correct?). The user had to score them on a numerical scale (San-Segundo et al, 2011). A subjective evaluation of sign language involves two main aspects: intelligibility and naturalness; both aspects influence the user's answer when general questions are included in the questionnaire. In order to isolate the intelligibility (the first target of this kind of system), the questionnaires were redesigned to avoid this aspect: the deaf customers where asked specific questions (instead of general ones) about some dialogues (for example, which bus you need to take to reach the Sanchinarro Hospital?). Three or four questions were considered per dialogue.

In order to improve the speech recognition rate, the speech recognizer was adapted to the bus company employees involved in the evaluation. For this adaptation, it was necessary to record 50 spoken sentences (1-2 sec.) for each employee.

The deaf customer's ages ranged from 31 to 60 years old with an average age of 43.4 years. Most of the Deaf customers said that they used a computer every day (6 Deaf users) or every week (2 Deaf users), the other two said never. Only half of them (5 Deaf users) had a medium-high understanding level of written Spanish and the other half had a low or very low level of understanding Spanish.

The evaluation of the speech into LSE translation module includes objective measurements from the system and subjective information. A summary of the objective measurements obtained from the system are shown in Table 5.

AGENT	MEASUREMENT	VALUE
System	Word Error Rate	5.9%
	Sign Error Rate (after translation)	9.4%
	Average Recognition Time	3.3 sec
	Average Translation Time	0.002 sec
	Average Signing Time	5.2 sec
	% of cases using example-based translation	96.3%
	% of cases using statistical translation	3.7%
	% of turns translating from speech recognition	94.0%
	% of turns translating from text	0.0%
	% of turns translating from text for repetition	6.0%
	# of bus company employee turns per dialogue	7.5
	# of dialogues	50
	Overall Time for translating Speech into LSE	8.5 sec

Table 5. Objective measurements for evaluating the Spanish into LSE translation system

The WER (Word Error Rate) for the speech recognizer is 5.9% being small enough to guarantee a low SER (Sign Error Rate) in the translation output: 9.4%. On the other hand, the time needed for translating speech into LSE (speech recognition + translation + signing) is around 8 seconds allowing a dialogue in real-time. Table 6 presents an analysis of the translation errors (9.4% in total) including an error classification, main causes and impact on the system.

Error description	Percentage	Main causes	Impact
Changes in the sentence structure and substitutions	**5.5%**	Problems in the sign sentence structure are mainly due to errors in the translation technology, selecting a wrong example or when dealing with sentence structures not seen in the collected corpus.	In these cases, the impact is the worst. The Deaf user does not understand anything and the bus company employee must repeat the information in a different way.

		These two kinds of errors have their main cause in speech recognition errors: insertions and deletions. Deletions are more frequent when the bus company employee lowers her/his voice, and they appear at the	Insertions have a negative impact. Sometimes, the Deaf user understood the Sign Language sentence but in many cases (>75%) the bus company employee had to repeat it.
Insertions	**1.9%**		
Deletions	**2.0%**	end of the sentence. Insertions appear when the government employee introduces additional noises into the speech (coughs, breathing, filled pauses "ehmm").	This is the error with the lowest impact. In many cases (>85%), the Deaf customer understood the overall meaning without repetition.

Table 6. Analysis of the errors generated by the translation system

Regarding to the questionnaires, Table 7 summaries the human recognition accuracy for isolated signs and for questions about the dialogues, based on the number of attempts (number of times the sentences are presented to the customer). The recognition accuracy includes all the experiments with the 10 deaf customers (from both sessions).

Human recognition rate depending on the number of times the LSE sentence was represented			
	1st	2nd	3rd
Isolated signs: 100 signs in total	80.0%	94.0%	100.0%
Questions about the dialogues: 167 questions in total	72.5%	82.0%	100.0%

Table 7: Recognition accuracy based on the number of times the sentences were represented. Percentages represent accumulative recognition rates.

For isolated signs the recognition rate in the first attempt is higher than for the dialogues. When evaluating isolated signs, the deaf people do not have the context to disambiguate different meanings of the same sign. On the other hand, when evaluating specific questions about the dialogues, the customers have the entire context but any small error in any sign can generate confusion. The main problems related to the recognition of some signs were problems on the orientation of several signs. It is also fair to report that there were discrepancies between Deaf people as to the correctness of some signs (i.e. the "FOTO" (photo) sign, it is represented by moving the index finger from both hands or only from the right hand) or the specific sign used (i.e. using the "FECHA" (date) sign instead of "DÍA" (day) sign). These discrepancies are solved in the real LSE conversations with a facial expression (i.e. pronouncing a word). In spite of the effort invested in this work, this aspect must be improved in the avatar. The sign specification was made based on the dictionary generated by Fundación CNSE, DILSE III. These discrepancies showed the need to keep working on the documentation process of the LSE. LSE is a young language with many variations in the different regions of Spain. Fundación CNSE (Confederación de Personas Sordas) is the national confederation including all local associations; FCNSE is making a significant effort to collect and document all of these variations. With this documentation, a Deaf user can learn these variations improving the communication between Deaf people coming from different regions in Spain. In the future, if LSE is included in TV subtitles, TV could reduce these discrepancies as has happened to other minority languages in Spain. Another source of discrepancy is the structure of some sign sentences. LSE, as in other languages, offers a high level of flexibility. This flexibility is sometimes not well understood and some of the possibilities are considered as wrong sentences. Some examples are presented in Table 8:

For the question "¿qué desea?" (What do you want?), the translation can be "QUERER QUÉ?" or "TU QUERER?" The system used the first one but some users preferred the second one.
Regarding the sign "MAPA" (map), some of the users think that it must go with the sign "CIUDAD" (city) in order to complement the meaning.

Table 8. Examples of discrepancy in sentence structure

Finally, some objective measurements of the spoken Spanish generation module are included in Table 9. These measurements have been obtained using a capturing software (Camtasia Studio 6: http://camtasia-studio.softonic.com/) and a detailed log generated by the system.

AGENT	MEASUREMENT	VALUE
System	Translation rate	96.5%
	Average translation time	0.001 sec
	Average time for text to speech conversion	2.1 sec
	% of cases using example-based translation	98.2%
	% of cases using statistical translation	1.8%
	Time for gloss sequence specification.	15.0 sec
	# of clicks for gloss sequence specification.	7.8 clicks
	# of glosses per turn	2.5
	% of utility use: - List of glosses - List of proposed next glosses	58.0% 42.0%
	# of turns using the most frequent sign sequences per dialogue	0.6
	# of deaf customer turns per dialogue	3.4
	# of dialogues	50
	Overall Time for translating LSE into Speech	17.1 sec

Table 9. Objective measurements for evaluating the Spanish generator from LSE

As is shown, the good translation rate and the short translation time make it possible to use this system in real conditions. Regarding the translation process, the example-based strategy has been selected in most of the cases. The parallel corpus generated is very good representative corpus for this kind of dialogue.

The user needed less than 20 seconds to specify a gloss sequence using the interface. This time is short considering that the deaf customer had only few minutes to practice with the visual interface before the evaluation. With more time for practicing, this time would be reduced.

In order to expand this analysis, Table 10 shows Spearman's correlation between some objective measurements from the Deaf customer evaluation and their background and age: computer experience, confidence with written Spanish, and age. This table also includes p-values for reporting the correlation significance. Because of the very low number of data and the unknown data distribution, Spearman's correlation has been used. This correlation produces a number between -1 (opposite behaviours) and 1 (similar behaviours). A 0 correlation means no relation between these two aspects.

EVALUATION MEASUREMENT	Computer experience	Confidence with written Spanish	Age
Questions answered at the first time (Table 7)	**0,52 (p=0.050)**	0,40 (p=0.114)	**-0,62 (p=0.040)**
Time for gloss sequence specification in the LSESpeak system (Table 9).	-0,35 (p=0.123)	-0,23 (p=0.214)	**0,58 (p=0.047)**
Percentage of times the bus company employee had to repeat a utterance (Table 5)	-0,26 (p=0.245)	-0,32 (p=0.122)	0,49 (p=0.056)

Table 10. Analysis of correlations between Deaf customer evaluation and their background

As is shown, only those results in bold are significant (p<0.05): the questions answered at the first time (Table 7) is positively correlated with the computer experience and negatively with age. Additionally, the time for gloss sequence specification (by the Deaf customer, Table 9) correlates positively with age.

Finally, comment that although the number of customers is not high enough to validate a complex engineering application like this, the field evaluation carried out has an important and interesting value to demonstrate the feasibility of a first version of the system. Evaluation with Deaf customers is very expensive, there are many people involved (researchers, bus company employees, deaf customers and interpreters) during several hours. Also, this description presents a proposal for performing this kind of evaluation, and reports initial experiments to compare with: if new researchers want to develop similar systems in other Sign Languages.

7. Conclusions

This paper has presented two systems for translating bus information into Spanish Sign Language (LSE: Lengua de Signos Española). One of the main contributions has been the analysis of different translation strategies and

their integration for obtaining the best accuracy: an example-based strategy, and a statistical translator. The translation module that integrates these two translation strategies has been used for developing two applications: the first for translating text messages from an information panel, and the second, for translating spoken Spanish from natural conversations at the information point of the bus company.

In the field evaluation of both systems, the sign language intelligibility is rather high although there are problems with the design of several signs: some problems are related to mistakes made by the research team, but other problems come about from the lack of normalization of the LSE. As regards the naturalness, it is true that the avatar signs the same sign in the same way, but this aspect is useful in making deaf people get used to the avatar. The more the system is used, the better the avatar is understood.

As regards the translation technologies, the whole translation presents an SER (Sign Error Rate) of less than 10% and a BLEU greater than 90%. This performance is very good but the overall times for translating Speech into LSE (8.5 seconds in Table 5) and LSE into Speech (17.1 seconds in Table 9) are high, providing a slow face-to-face interaction (3-5 minutes) compared to a speech-speech interaction (less than 1 minute). Anyway, the field evaluation shows the interest of considering these translation technologies in real applications. These applications are especially interesting when a human interpreter is not available.

Acknowledgements

References

Adamo-Villani N. "3D Rendering of American Sign Language finger spelling: a comparative study of two animation techniques" 5th International Conference on Computer Instructional Technologies, Italy, pp. 808-812, 2008.

Agarwal, Abhaya and Lavie, Alon, 2008. "Meteor, m-bleu and m-ter: Evaluation Metrics for High-Correlation with Human Rankings of Machine Translation Output", Proceedings of Workshop on Statistical Machine Translation at the 46th Annual Meeting of the Association of Computational Linguistics (ACL-2008), Columbus, June 2008.

Banerjee, S. and A. Lavie, 2005. "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments", Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43th Annual Meeting of the Association of Computational Linguistics (ACL-2005), Ann Arbor, Michigan, June 2005.

Bungeroth J., Ney, H.,: Statistical Sign Language Translation. In Workshop on Representation and Processing of Sign Languages, LREC 2004, 105-108.

Casacuberta F., E. Vidal. 2004. "Machine Translation with Inferred Stochastic Finite-State Transducers". Computational Linguistics, Vol. 30, No. 2, pp. 205-225, June 2004.

Cemil Oz, Ming C. Leu. 2011 "American Sign Language word recognition with a sensory glove using artificial neural networks" Engineering Applications of Artificial Intelligence, Volume 24, Issue 7, October 2011, Pages 1204-1213

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.

Gonzalez-Morcillo C., Weiss G., Vallejo D., Jimenez L. and Castro-Schez J.J. "A MultiAgent Architecture for 3D Rendering Optimization", Applied Artificial Intelligence, Vol 24, No 4, 313-349, 2010.

Herrera V., Gonzalez-Morcillo C., Garcia M.A., Mateos J.A., Arriaga I. "Ganas: A Flexible Architecture for 3D Sign Language Rendering", International Conference Interfaces Human and Computer Interaction, pp. 61-68, Portugal, 2009.

Ibarguren, A., Maurtua, I., Sierra. B., 2010. "Layered architecture for real time sign recognition: Hand gesture and movement" Engineering Applications of Artificial Intelligence, Volume 23, Issue 7, October 2010, Pages 1216-1228

Kipp M., Heloir A. Nguyen Q. "Sign language avatars: Animation and comprehensibility", Intelligent Virtual Agents, pp. 113-126, Springer, 2011.

Koehn P., F.J. Och D. Marcu. 2003. "Statistical Phrase-based translation". Human Language Technology Conference 2003 (HLT-NAACL 2003), Edmonton, Canada, pp. 127-133, May 2003.

Levenshtein. V., 1966. "Binary codes capable of correcting deletions, insertions, and reversals". Soviet Physics Doklady, 1966.

Lever N. "Real Time 3D Character Animation with Visual C++", Focal Press, 2002.

López-Ludeña, V., San-Segundo, R., Montero, J.M., Córdoba, R., Ferreiros, J., Pardo, J.M., "Automatic Categorization for Improving Spanish into Spanish Sign Language Machine Translation" in press Computer Speech and Language, 2011.

López-Ludeña, V., San-Segundo, R., González Morcillo, C., López, J.C., Pardo, J.M.,. 2013a. "Increasing adaptability of a speech into sign language translation system". Expert Systems With Applications. Volume 40, Issue 4, March 2013a, Pages 1312–1322.

López-Ludeña, V., Barra-Chicote, R., Lutfi, S., Montero, J.M., San-Segundo, R.. 2013a. "LSESpeak: A spoken language generator for Deaf people". Expert Systems With Applications. Volume 40, Issue 4, March 2013b, Pages 1283–1295.

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, n° 4, pp. 527-549.

Morrissey S., and Way A.,. 2005. "An example-based approach to translating sign language". In Workshop Example-Based Machine Translation (MT X–05), pages109–116, Phuket, Thailand, September.

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.

Niewiadomski R., Bevacqua E., Mancini M., Pelachaud C. "Greta: an interactive expressive ECA system" 8th International Conference on Autonomous Agents and Multiagent Systems, Vol. 2, pp 1399-1400, 2009.

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, October 2000.

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.

Och J., Ney. H., 2003. "A systematic comparison of various alignment models". Computational Linguistics, Vol. 29, No. 1 pp. 19-51, 2003.

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.

Pfau R., Quer J. "Nonmanuals: their gramatical and prosocid roles" Sign languages, pp 381-402, Cambridge University Press, 2010.

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.

San-Segundo, R., Montero, J.M., Córdoba, R., Sama, V., Fernández, F., D'Haro, L.F., López-Ludeña, V., Sánchez D., and García, A., 2011. Design, development and field evaluation of a Spanish into sign language translation system. Pattern Analysis and Applications. Pattern Analysis and Applications. Volume 15, Issue 2, pp 203-224.

Stolcke A. "SRILM – An Extensible Language Modelling Toolkit". ICSLP. 2002. Denver Colorado, USA.

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.

Thiebaux M., Marsella S., Marshall A.N., Kallman M. "Smartbody: Behavior realization for embodied conversational agents" 7th International Joint Conference on Autonomous Agents and Multiagent Systems, Vol 1, pp 151-158, 2008.

Zens R., F.J. Och, H. Ney. 2002. "Phrase-Based Statistical Machine Translation". German Conference on Artificial Intelligence (KI 2002). Aachen, Germany, Springer, LNAI, pp. 18-32, Sep. 2002.

Zwiterslood, I., Verlinden, M., Ros, J., van der Schoot, S., 2004. "Synthetic Signing for the Deaf: eSIGN" .Proceedings of the Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment, CVHI 2004, 29 June-2 July 2004, Granada, Spain.

Watt A., Watt M. "Advanced Animation and Rendering Techniques", Pearson Education Harlow, UK, 1992.