Expert Systems with Applications 40 (2013) 1312–1322

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Increasing adaptability of a speech into sign language translation system

Verónica López-Ludeña [a,*], Rubén San-Segundo [a], Carlos González Morcillo [b], Juan Carlos López [b], José M. Pardo Muñoz [a]

[a] Grupo de Tecnología del Habla, Departamento de Ingeniería Electrónica, ETSI Telecomunicación, Universidad Politécnica de Madrid, Spain
[b] Grupo de Sistemas Inteligentes Aplicados, Departamento de Tecnologías y Sistemas de Información, Universidad de Castilla-La Mancha, Spain

ARTICLE INFO

Keywords:
Adaptation
New domain
Few resources
Deaf people
Spanish sign language (LSE)
Spoken language translation
Sign animation

ABSTRACT

This paper describes a new version of a speech into sign language translation system with new tools and characteristics for increasing its adaptability to a new task or a new semantic domain. This system is made up of a speech recognizer (for decoding the spoken utterance into a word sequence), a natural language translator (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). In order to increase the system adaptability, this paper presents new improvements in all the three main modules for generating automatically the task dependent information from a parallel corpus: automatic generation of Spanish variants when generating the vocabulary and language model for the speech recogniser, an acoustic adaptation module for the speech recogniser, data-oriented language and translation models for the machine translator and a list of signs to design. The avatar animation module includes a new editor for rapidly design of the required signs. These developments have been necessary to reduce the effort when adapting a Spanish into Spanish sign language (LSE: Lengua de Signos Española) translation system to a new domain. The whole translation presents a SER (Sign Error Rate) lower than 10% and a BLEU higher than 90% while the effort for adapting the system to a new domain has been reduced more than 50%.

1. Introduction

Deafness is the impairment that worst affect the communication between people. Deaf people (especially for 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 became apparent because the lack of feedback in speak-listen procedure. This fact causes deaf people to have problems when accessing information, education, job, social relationship, culture, etc. According to information from INE (Statistic Spanish Institute), in Spain, there are 1,064,000 deaf people. 47% of deaf population do not have basic studies or illiterate, and only between 1% and 3% have finished their university studies (as opposed to 21% of Spanish hearing people). Another example are the figures from the National Deaf Children's Society (NDCS), Cymru, revealing for the first time a shocking attainment gap between deaf and hearing pupils in Wales. In 2008, deaf pupils were 30% less likely than hearing pupils to gain five A*-C grades at General Certificate of Secondary Education (GCSE) level, while at key stage 3 only 42% of deaf pupils achieved the core subject indicators, compared to 71% of their hearing counterparts. Another example is a study carried out in Ireland in 2006; of 330 respondents "38% said they did not feel confident to read a newspaper and more than half were not fully confident in writing a letter or filling out a form" (Conroy, 2006).

However, Deaf[1] people 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 to people with learning problems but this is not true. 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. Sign languages are not disseminated enough among hearing people appearing communication barriers between Deaf and hearing people. These barriers are even more problematic when they appear between a deaf person and a government

* Corresponding author. Address: Grupo de Tecnología del Habla, Departamento de Ingeniería Electrónica, ETSI Telecomunicación, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain.
E-mail address: veronicalopez@die.upm.es (V. López-Ludeña).

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

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.08.059

employee who is providing a face-to-face service, for example. These barriers reveal the need of developing new technologies for improving the communication between hearing and Deaf people. A good example is the development of speech into sign language translation systems.

The main problem of sign languages is the reduced amount of economical resources available for minority languages. This lack of economical resources is the main cause of the reduced number of interpreters. In the USA, there are 650,000 Deaf people (who use a sign language) but there are only 7000 sign-language interpreters, i.e. a ratio of 93 deaf people to 1 interpreter. In Finland, there is the best ratio, 6–1, and in Slovakia the worst with 3000 users to 1 interpreter (Wheatley and Pabsch, 2010). In Spain, this ratio is 221–1. Another effect is the limited amount of language resources for researching in speech and language technologies involving sign languages. One important indicator reflecting this aspect is the low number of sign languages corpora and their small size (San-Segundo et al., 2010). When developing natural language interfaces, a significant amount of resources is required to model task knowledge properly. For the case of language translation systems, it is necessary a parallel corpus including source and target language sentences. In the literature, there are small parallel corpora involving sign languages, increasing the difficulty of developing speech into sign language translation systems.

This paper describes a new version of a Spanish into Spanish sign language (LSE: Lengua de Signos Española) translation system (San-Segundo et al., 2012) with new tools and characteristics for increasing its adaptability to a new task or a new semantic domain. These tools will permit to adapt the translation system to a new domain with a limited parallel corpus. The working time, necessary to adapt the system, is reduced significantly. This paper presents improvements and new characteristics in all the modules composing the speech into sign language translation system: speech recognizer, language translator and sign representation interface. This research work is included in the CONSIGNOS project (Plan Avanza Exp N: TSI-020100-2010-489). This project aims to adapt a Spanish into Spanish sign language (LSE: Lengua de Signos Española) translation system, already developed to a specific domain (like driver's licence renewing service), to different domains: travel information and hotel reception.

This paper is organised as follows. Section 2 presents the state of the art. Section 3 describes the system architecture including speech recognition, language translation and sign animation modules. Section 4 describes the speech recognition system and the new features introduced for improving its adaptability. Section 5 presents the new characteristics of the language translation module. Section 6 includes the new module for sign animation using an animated character and the sign editor developed for rapidly sign prototyping. Finally, Section 7 includes the main conclusions of this work.

2. State of the art

Machine translation has been one of the main research topics funded by the European Commission (EC) and the USA Government in speech and language processing. In Europe, TC-STAR has been one of the most important projects. The TC-STAR project (http://www.tc-star.org/) was envisaged as a long-term effort to advance research into all core technologies for Speech-to-Speech Translation (SST). Another important project on language translation funded by the EC is EuroMatrixPlus (http://www.euromatrixplus.net/). This project focuses on creating example systems for every official EU language, and providing other machine translation developers with a baseline infrastructure for building statistical translation models. The EuroMatrixPlus team has organized several Workshops on Statistical Machine Translation (SMT). On the webpages http://www.statmt.org/ and http://matrix.statmt.org/, it is possible to obtain all the information about these events. As a result of these workshops, there is a free machine translation system called Moses and available from this web page (http://www.statmt.org/moses/). Moses is a phrase-based statistical machine translation system that allows machine translation system models to be built for any pair of languages, using a parallel corpus.

In the USA, DARPA (Defence 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. Automatic processing "engines" convert and distil the data, delivering pertinent, consolidated information in easy-to-understand formats to military personnel and monolingual English-speaking analysts in response to direct or implicit requests. GALE consists of three major engines: Transcription, Translation and Distillation. The output of each engine is English text. The input to the transcription engine is speech and to the translation engine, text. This project has been active for two years, and the GALE contractors have been engaged in developing highly robust speech recognition, machine translation, and information delivery systems in Chinese and Arabic. This program has also been boosted by the machine translation evaluation organised by the USA Government, NIST (National Institute of Standards and Technology) (http://www.itl.nist.gov/iad/mig/tests/mt/).

In recent years, several groups have developed prototypes for translating spoken language into sign language: example-based (Morrissey, 2008), rule-based (Marshall & Sáfár, 2005; San-Segundo et al., 2008), full sentence (Cox et al., 2002), statistical (Morrissey, Way, Stein, Bungeroth, & Ney, 2007; Stein, Bungeroth, & Ney, 2006; Vendrame & Tiotto, 2010), or hierarchical approaches (San-Segundo et al., 2012). All of these translation systems are focused on specific semantic domains and they have been developed using small corpora. The effort for developing every of these systems has been very important and, in many cases, it has been necessary the intervention of experts for developing translation rules or including source language variants. This kind of systems can complement a Sign Language into Speech translation system, allowing a two direction interaction. In (Karami, Zanj, & Sarkaleh, 2011) a system for recognizing static gestures of alphabets in Persian sign language (PSL) using Wavelet transform and neural networks is presented. A system for automatic translation of static gestures of alphabets and signs in American Sign Language is presented by using Hough transform and neural networks trained to recognize signs in Munib, Habeeb, Takruri, and Al-Malik (2007). In Sylvie and Surendra (2005) a review of research into sign language and gesture recognition is presented.

In the last five years, several projects have started to generate more parallel corpora including sign languages: in American Sign Language (Dreuw, Neidle, Athitsos, Sclaroff, & Ney, 2008), British Sign Language (Schembri, 2008), Greek Sign Language (Efthimiou & Fotinea, 2008), in Irish Sign Language (Morrissey, Somers, Smith, Gilchrist, & Dandapat, 2010), NGS (German Sign Language) (Hanke, König, Wagner, & Matthes, 2010), Italian Sign Language (Geraci et al., 2010) and Spanish sign language (LSE: Lengua de Signos Española) (San-Segundo et al., 2010). But this is not enough for minority languages with low resources. Researchers must invest more effort on developing new tools and features to improve technology adaptability. This adaptability is defined as the capability of reducing significantly the effort and the parallel corpus needed for adapting a speech into sign language translation system to a new domain. This is the main focused of the research work described in this paper.

Signing avatars are a relatively novel research field. In the last 15 years some relevant results have been obtained. Almost every sign language representation system uses the gloss notation where each entity corresponds to one sign. Nevertheless in this notation the gloss does not describe how to represent the sign. In the research area of signing avatars, there are two main approaches: articulatory, that generates synthetic animations on the fly based on a motion specification language and concatenate which uses a pre-recorded videos of human motion (Huenerfauth & Hanson, 2009). Two of the most relevant projects ViSiCAST and eSIGN (Elliott, Glauert, Kennaway, Marshall, & Safar, 2008) were developed based on the Hamburg Notation System for Sign Language HamNoSyS (Prillwitz, Leven, Zienert, Hanke, & Henning, 1989). A new technology called Animgen was developed within these projects, where an avatar (e.g. Guido was one of the most relevant avatar for sign language) represents an XML version of the HamNoSyS language called SiGML. The main drawback of Animgen is the unfeasibility to modify the resulting animations. To overcome these limitations, there are recent proposals such as Zedebee that allows parametrizable scripts (Filhol, 2009) or Paula which incorporates nonverbal components and natural pose calculation (Wolfe, McDonald, Davidson, & Frank, 2007). Another relevant project developed at the University of Tunis is Websign (Jaballah & Jemni, 2010), that is mainly based on web technologies and makes use of a virtual avatar.

There are also some companies which are developing commercial sign animation systems, such as Vcom3D (http://www.vcom3d.com/) that allows edition and creation of new signs (using the GestureBuilding tool). IBM is also investing resources in the development of SiSi (Paulson, 2008); an application which uses ViaVoice to recognise the language and use a dictionary of signs to generate the final animation. Ohali provides a comparison of some of the most relevant commercial avatars developed in this field (Ohali, 2010).

3. Speech into sign language overview

This section introduces the system overview and the new tools and utilities for adapting the system to a new domain.

Fig. 1 shows the module diagram developed for translating spoken language into LSE. The first module, the automatic speech recognizer (ASR), converts natural speech into a sequence of words (text). It uses a dictionary, a language model and acoustic models for every allophone. The natural language translation module converts a word sequence into a sign sequence. For this module, the paper presents and combines two different strategies. The first one 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 one 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. 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 the use in different devices.

At the bottom part of Fig. 1, the new tools for increasing the system flexibility are presented. The speech recogniser includes an acoustic adaptation module, for adapting the acoustic models to a new specific environment (indoor or outdoor scenarios), a new speaker, or a new Spanish accent. The speech recogniser uses a vocabulary and a language model generated from the parallel corpus (only source language) automatically. In this process, a new module has been included for introducing source language variants, increasing the speech recogniser flexibility. The language translation module presents a new configuration compared to the previous version (San-Segundo et al., 2012) where all the translation strategies are data-oriented ones. With this new design, the required models are generated automatically from a parallel corpus. The statistical translation strategy incorporates a new pre-processing module (López-Ludeña et al., 2012) that permits to increase its performance. This increment has allowed replacing a rule-based translation strategy with a statistical one. The sign animation module includes a new version of the sign editor that incorporates new options (like predefined positions and orientations) for reducing significantly the sign specification time.

4. Automatic speech recogniser

The speech recognizer used is a state of the art speech recognition system developed at Speech Technology Group (GTH-UPM:

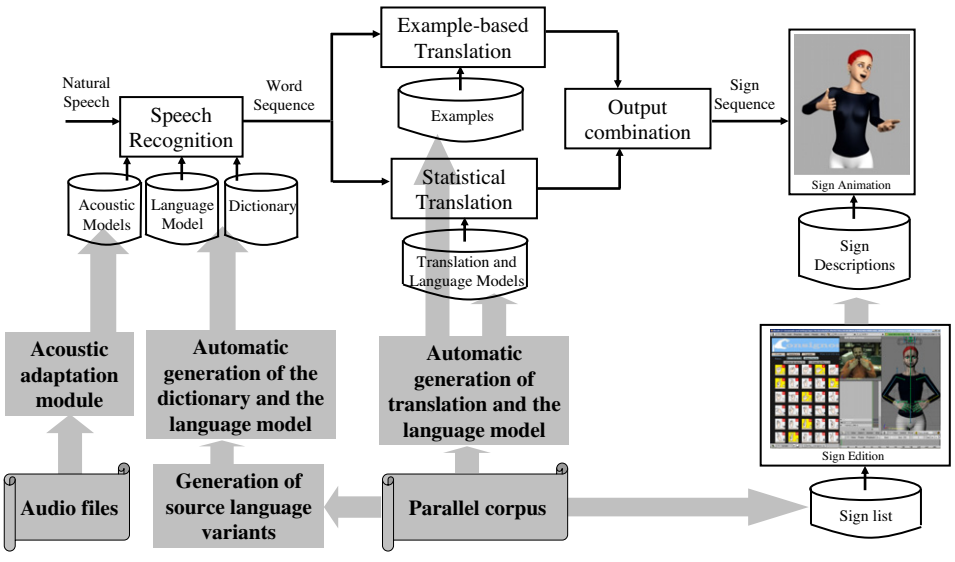

Fig. 1. Diagram of the speech into sign language translation system including new tools and characteristics for increasing its adaptability.

http://lorien.die.upm.es). It is an HMM (Hidden Markov Model)-based system able to recognize continuous speech: it recognizes utterances formed by several words continuously spoken. In this application, the vocabulary size is 653 Spanish words: the corpus vocabulary (with 527 words) was extended with a complete list of numbers (from 0 to 100), weekdays, months, etc. (Fig. 2).

The recognizer has been trained with a lot of speakers (4000 people), making it robust against a great range of potential speakers without further training by actual users. The recognizer has been trained by using more than 40 h of speech from the SpeechDat database (Moreno, 1997). The system uses a front-end with Perceptual Linear Predictive (PLP) (Hermansky, 1990) coefficients derived from a Mel-scale filter bank (MF–PLP). This front-end includes Cepstral Mean Normalization (CMN) and Cepstral Variance Normalization (CVN) techniques (Jankowski, Hoang-Doan, & Lippmann, 1995).

For Spanish, the speech recognizer uses a set of 45 units: it differentiates between stressed/unstressed/nasalized vowels, it includes different variants for the vibrant 'r' in Spanish, different units for the diphthongs, the fricative version of 'b', 'd', 'g', and the affricates version of 'y' (like 'ayer' and 'cónyuge'). The system also has 16 silence and noise models for detecting acoustic sounds (non speech events like background noise, speaker artefacts, filled pauses, etc.) that appear in spontaneous speech. The system uses context-dependent continuous Hidden Markov Models (HMMs) built using decision-tree state clustering: 1807 states and 7 mixture components per state.

About the language model, the recognition module uses statistical language modelling: 2-grams, as the database is not large enough to estimate reliable 3-grams.

The recognition system can generate one optimal word sequence (given the acoustic and language models), a solution expressed as a directed acyclic graph of words that may compile different alternatives, or even the N-best word sequences sorted by similarity to the spoken utterance. In this work, only the optimal word sequence is considered.

The recognizer provides one confidence measure for each word recognized in the word sequence. The confidence measure is a value between 0.0 (lowest confidence) and 1.0 (highest confidence) (Ferreiros, San-Segundo, D'Haro, & Barra, 2005). This measure is important because the speech recognizer performance varies depending on several aspects: level of noise in the environment, non-native speakers, more or less spontaneous speech, or the acoustic similarity between different words contained in the vocabulary.

As regards the performance of the speech recogniser module, with vocabularies of less than 1000 words, the Word Error Rate (WER) is less than 5%.

In this work, three new utilities have been incorporated to increase its flexibility:

- The first one consists of an acoustic adaptation module. The acoustic models can be adapted to one speaker or to a specific acoustic environment using the Maximum a Posteriori (MAP) technique (Gauvain & Lee, 1994). This module has been very important for adapting the acoustic model to new different acoustic environments: the reception of a hotel and in a small cabin situated in the street. Other interesting possibility has been to adapt the acoustic model to a specific speaker. In this situation, the WER is reduced significantly (less than 3%) and the system speed increased very much (more than 50%).
- One important problem of the speech recogniser when generating its dictionary and language model from a small corpus is the high number of Out of Vocabulary words (OOVs) and the poor estimation of the language model probabilities. In order to deal with this problem, in this work a new module of adding Spanish variants has been incorporated. This modules includes, automatically, several variants considering this aspects:
  - The system introduces variants for formal and no-formal ways of referring to "you" ("usted" or "tu" in Spanish). For example, given the colloquial form "tu debes darme una foto" ("you must give me a photo"), the system would include "usted debe darme una foto" (with the same translation in English "you must give me a photo" and also in LSE).
  - Including synonymous for some names, adjectives and verbs. The system detects synonymous in different sentences and generates variants interchanging these synonymous words.
  - Changing the order of expressions like "please" or "thank you": "¿Podrías decirme dónde coger el autobús 42?, por favor" ->"Por favor, ¿podrías decirme dónde coger el autobús 42?" ("Please, Could tell me where can I take the 42 bus?")
- The language model is based on classes. Instead of considering individual words for estimating the n-gram sequence probabilities, the system train probabilities of word and class sequences. Every class can contain several words. This utility is very interesting when, for example, weekdays or months appear in the domain. With small corpora, there are not enough sentences including all possible weekdays or months. Including these words in classes (WEEKDAYS or MONTHS) help to train better the language model (Brown et al., 1990).

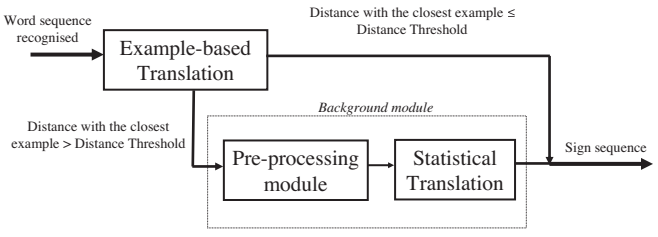

Fig. 2. Diagram of natural language translation module combining two different translation strategies.

5. Language translation system

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 in order to look for the best possible match. If the distance with the closest example is lower than a threshold (Distance Threshold), the translation output is the same as the example translation. But if the distance is higher, a background module based on a statistical strategy translates the word sequence. During the developing tests, the best results were obtained for a Distance Threshold (DT) ranging from between 20% and 30%. In the field evaluation, the DT was fixed at 30% (one difference is permitted in a 4-word sentence).

The main contribution in the language translator has been focused on the background module. The statistical translation strategy incorporates a new pre-processing module (López-Ludeña et al., 2012) that permits to increase its performance. This increment allowed replacing a rule-based translation strategy (presented in the previous version of the system (San-Segundo et al., 2012)) with a statistical one. Thanks to this replacement, it is not necessary to generate any rule by hand, avoiding any manual intervention for developing the language translation module from the parallel corpus. The statistical translation module is based on Moses, an open-source phrase-based translation system released from NAACL Workshops on Statistical Machine Translation (http://www.statmt.org) in 2011.

In the next sections, more details about the translation strategies will be presented.

5.1. Example-based translation strategy

An example-based translation system uses a set of sentences in the source language and their corresponding translations in the target language, for translating other similar source-language sentences. In order to determine whether one example is equivalent (or at least, similar enough) to the sentence to be translated, the system computes a heuristic distance between them. By defining a threshold on this heuristic distance, it is possible to define how similar the example must be to the sentence to be translated, in order to consider that they generate the same target sentence. If the distance is lower than a threshold, the translation output will be the same as the example translation. But if the distance is higher, the system cannot generate any output. Under these circumstances, it is necessary to consider other translation strategies.

The heuristic distance used in the first version of the system was a modification of the well-known Levenshtein distance (LD) (Levenshtein, 1966). The heuristic distance is the LD divided by the number of words in the sentence to be translated (this distance is represented as a percentage). One problem of this distance is that two synonyms are considered as different words (a substitution in the LD) while the translation output can be the same. In recent work (San-Segundo et al., 2012), the system has been modified to use an improved distance where the substitution cost (instead of being 1 for all cases) ranges from 0 to 1 depending on the translation behaviours of the two words. Additionally, the deletion cost ranges also from 0 to 1 depending on the probability of not aligning a word to any sign (this word is associated to the NULL tag). These behaviours are obtained from the lexical model computed in the statistical translation strategy (described in next section). For each word (in the source language), an N-dimension translation vector ($\hat{w}$) is obtained where the "i" component, $P_w(g_i)$, is the probability of translating the word "w" into the sign "s_i". N is the total number of signs (sign language) in the translation domain. The sum of all vector components must be 1. The substitution cost between words "w" and "u", and the deletion cost of word "w" are given by the following equation.

Substitution and deletion costs based on the translation behaviour

$$Subs.\,Cost(w,u) = \frac{1}{2}\sum_{i=1}^{N} abs(P_w(s_i) - P_u(s_i)) \qquad Del.Cost(w) = P_w(NULL) \tag{1}$$

When both words present the same behaviour (the same vectors), the probability subtraction tends towards 0. Otherwise, when there is no overlap between translations vectors, the sum of the probability subtractions (in absolute values) tends towards 2. Because of this, the ½ factor has been included to make the distance range from 0 to 1. These costs are computed automatically so it is not necessary any manual intervention to adapt the example-based translation module to on a new semantic domain using only a parallel corpus.

The biggest problem with an example-based translation system is that it needs large amounts of pre-translated text to make a reasonable translator. In order to make the examples more effective, it is possible to generalize them (Brown, 2000), so that more than one string can match the same example, increasing its flexibility.

5.2. Statistical translation strategy

The statistical translation module is composed of a pre-processing stage and a phrase-bases translation system.

5.3. Pre-processing module

This pre-processing module replaces Spanish words with associated tags (López-Ludeña et al., 2012). The pre-processing module uses a word-tag list for tagging the source sentence. In this module, all the words in the input sentence are replaced by their tags with the exception of those words that do not appear in the list (OOV words). They are kept as they are. After that, the "non-relevant" tags are removed from the input sentence (Non-relevant words are Spanish words not assigned to any sign). The word-tag list is generated automatically using the lexical model obtained from the word-sign GIZA++ alignments (Och & Ney, 2003). Given the lexical model, the tag associated to a given word is the sign with the highest probability of being the translation of this word. But this tag is assigned only if this probability is higher than a threshold. If there is no probability higher than the threshold, the tag for this word will be the same word. If the most probable sign is "NULL" and its probability is higher than this threshold, this word will tagged with the "non-relevant" tag. This probability threshold is fixed to 0.4 based on development evaluations. Generating these tags automatically is a very important aspect because it contributes to the main target of this work: reducing the time for adapting the translation system to a new semantic domain.

In conclusion, the pre-processing module allows the variability in the source language to be reduced together with the number of tokens that make up the input sentence. These two aspects give rise to a significant reduction in the number of source-target alignments the system has to train. When having a small corpus, as is the case in many sign languages, this reduction of alignment points permits to get better training models with fewer data.

5.4. Phrase-based translation module

The Phrase-based translation system is based on the software released at the 2011 NAACL Workshop on Statistical Machine Translation (http://www.statmt.org/wmt11/) (Fig. 3).

In this study, a phrase consists of a subsequence of words (in a sentence) that intends to have a meaning. Every sentence is split in several phrases automatically so this segmentation can have errors. But, the main target, when training a phrase-based model, is to split the sentence in several phrases and to find their corresponding translations in the target language.

The phrase model has been trained starting from a word alignment computed using GIZA++ (Och & Ney, 2003). GIZA++ is a statistical machine translation toolkit that is used to train IBM Models 1–5 and an HMM word alignment model. In this step, the alignments between words and signs in both directions (Spanish-LSE and LSE-Spanish) are calculated. The "alignment" parameter has been fixed to "target-source" as the best option (based on experiments over the development set): only this target-source alignment was considered (LSE-Spanish). In this configuration, alignment is guided by signs: this means that in every sentence pair alignment, each word can be aligned to one or several signs (but not the opposite), and also, it is possible that some words were not aligned to any sign. When combining the alignment points from all sentences pairs in the training set, it is possible to have all possible alignments: several words aligned to several signs.

After the word alignment, the system performs a phrase extraction process (Koehn, Och, & Marcu, 2003) where all phrase pairs that are consistent with the word alignment (target-source alignment in our case) are collected. In the phrase extraction, the maximum phrase length has been fixed at 7 consecutive words, based on development experiments over the development set (see previous section).

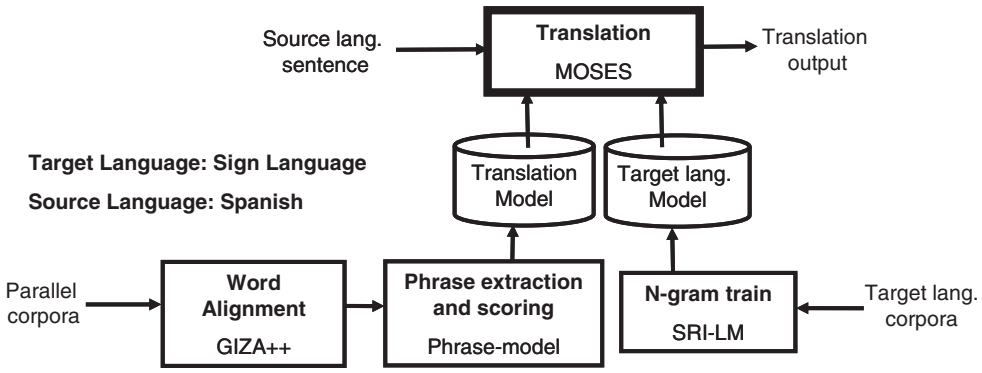

Fig. 3. Phrase-based translation architecture.

Finally, 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.

For the translation process, the Moses decoder has been used (Koehn, 2010). This program is a beam search decoder for phrase-based statistical machine translation models. In order to obtain a 3-grams language model, the SRI language modeling toolkit has been used (Stolcke, 2002).

5.5. Translation experiments

In order to evaluate the translation module, some experiments have been carried out using the whole Spanish-LSE parallel corpus described in San-Segundo et al. (2010). The corpus was divided randomly into three sets: training (75% of the sentences), development (12.5% of the sentences) and test (12.5% of the sentences), carrying out a Cross-Validation process. Table 1 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. Another metric is BLEU (BiLingual Evaluation Understudy; (Papineni, Roukos, Ward, & Zhu, 2002;), and finally, NIST (http://www.nist.gov/speech/tests/mt/). It is important to underline that SER and PER are error metrics (a lower value means a better result) while BLEU and NIST are accuracy metrics (a higher value means a better result).

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

Confidence Interval at 95%

$$\pm\Delta = 1,96\sqrt{\frac{SER\ \ (100 - SER)}{n}} \quad (2)$$

n is the number of signs used in testing, in this case n = 12,741. An improvement between two systems is statistically significant when there is no overlap between the confidence intervals of both systems. As is shown in Table 1, all improvements between different approaches are higher than the confidence intervals.

As is shown in Table 1, example-based and statistical strategies have SER greater than 20%. Important increment has been reached in the statistical approach compared to the previous system (San-Segundo et al., 2012) but a SER close to 20% is still a bit high. Table 1 also presents the translation results for the example-based approach for those sentences that have a heuristic distance (with the closest example) lower 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 1 presents the results for the combination of several translation strategies: example-based (considering a heuristic distance <30%) and statistical approaches. As is shown, with the hierarchical system it is possible to obtain better results by translating all the test sentences: SER <10%. Combining both translation strategies allows reaching a good compromise between performance and flexibility when the system is trained with a small parallel corpus.

6. 3D Avatar animation module

The systems described in the state of the art section are based on interactive rendering techniques. These techniques reduce the time spent in the rendering phase, but produce movements with a lower degree of realism than those obtained through non-interactive rendering techniques. Most of these projects also do not support the creation of new signs. The representation module of ConSignos has been designed to overcome the drawbacks of existing proposals. The animation module is focused on obtaining realistic movements by means of the combination of independent animation channels. The main goals of this approach are the comprehensibility of the generated sentences and the independence on the display device (web page, mobile phone, TV...).

6.1. General description

The animation module uses a declarative abstraction module used by all of the internal components. This module used a description based on XML, where each key pose configuration is stored defining its position, rotation, length and hierarchical

Table 1
Result summary for example-based, rule-based and statistical approaches.

		SER (%)	±Δ	PER (%)	BLEU	NIST
Example-based strategy	Reference	34.2	0.84	33.4	0.6212	7.554
	ASR output	39.3	0.85	37.3	0.5531	6.823
Example-based approach (considering a heuristic distance < 30%)	Reference	3.8	0.40	3.2	0.9512	10.752
	ASR output	6.8	0.44	5.2	0.9212	10.252
Statistical strategy	Reference	19.9	0.72	17.9	0.7827	9.233
	ASR output	24.8	0.75	22.7	0.7347	8.321
Combining translation Strategies	Reference	6.4	0.44	5.03	0.9211	10.232
	ASR output	7.8	0.47	6.65	0.9156	10.145

structure. We have used an approximation of the standard defined by H-Anim (Humanoid Working Group ISO/IEC FCD 19774:200×). In terms of the bones hierarchy, each animation chain is composed by several «joint» objects that define transformations from the root of the hierarchy (Fig. 4).

Several general purpose avatars such as Greta (Niewiadomski, Bevacqua, Mancini, & Pelachaud, 2009) or SmartBody (Thiebaux, Marsella, Marshall, & Kallman. M., 2008) lacked an important 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 constraints. This is the most time-consuming phase in the generation of a new sign and, as exposed in next section, a new approach to increase the adaptability has been created. For each sign it is necessary to model nonmanual features (torso movements, facial expressions and gaze). For the upper body control, some high-level IK control has been defined (see Fig. 5).

The skeleton defined in the representation module is composed by 103 bones, out of which 19 are inverse kinematics handlers (they have influence over a set of bones). The use of inverse kinematics and spherical quaternion interpolation (Watt & Watt, 1992) ease the work of the animators to capture 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 over 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 an articulation moves. In most kinematics configurations is essential to define rotation constraints to avoid forbidden configurations and simulate only physically correct positions. There are two ways to deal with IK: analitic or iterative

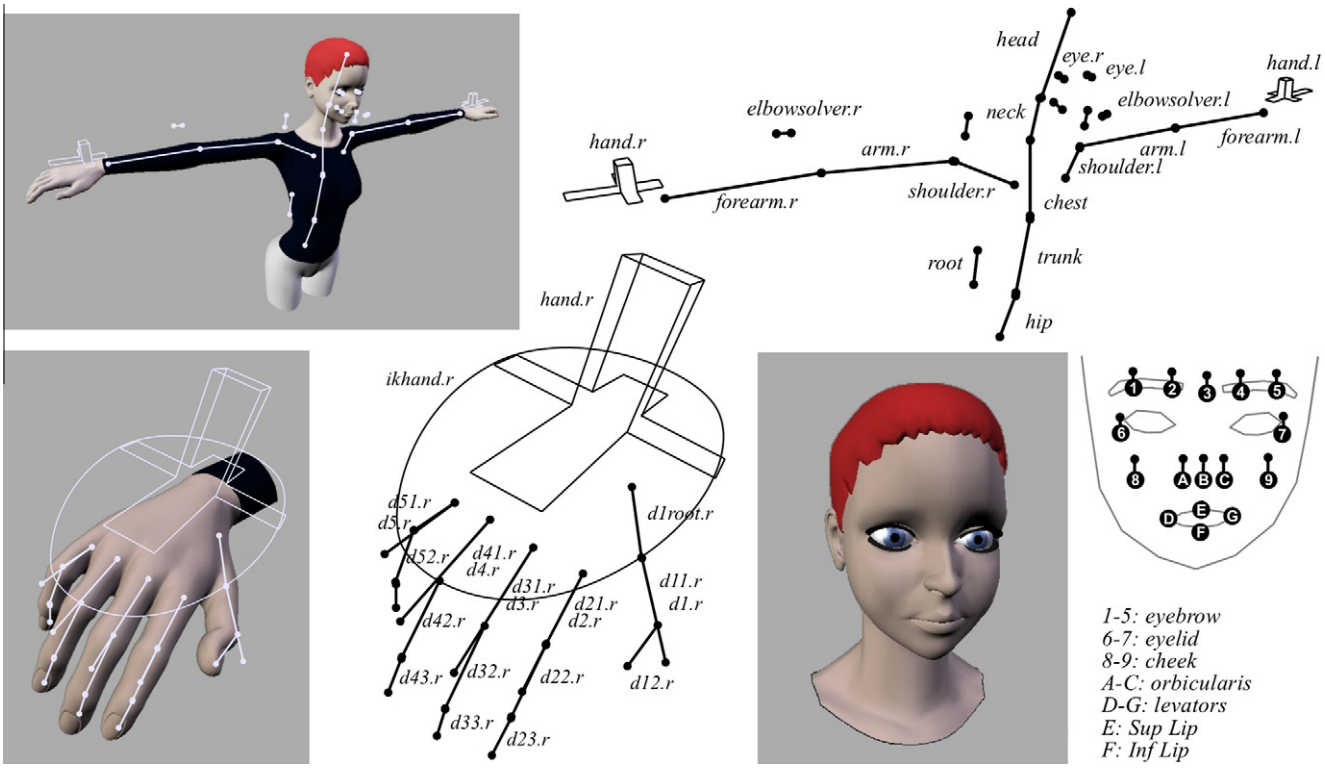

Fig. 4. Main bones and Inverse Kinematics controls (body, hand and face) of the avatar.

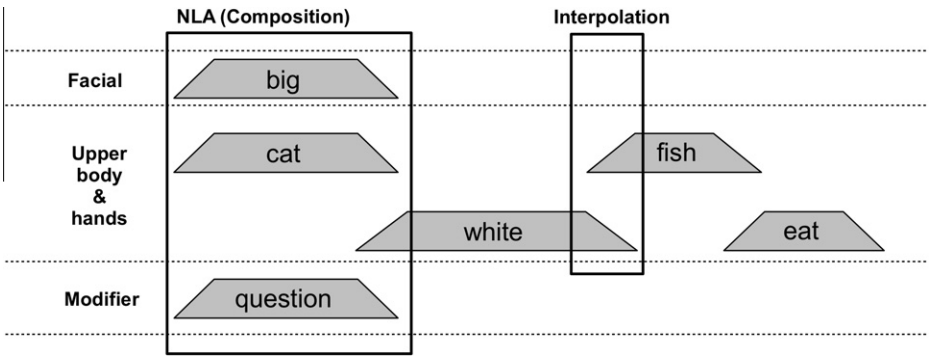

Fig. 5. Composition of animation channels (NLA) and interpolation between signs.

methods. The analitic methods require 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, ConSignos 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 over the elements, adjusting the configuration of each joint until the position of the effector is close to the desired position or a concrete 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 nonmanual animation is used to remark adjectival or adverbial information. The movements of the mouth are also highly important to focus the visual attention making the comprehension easier. As pointed out by Pfau and Quer (2010), nonmanuals 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 are used in film production to merge individual actions into complex animations. Each small piece of the animation (action) is specialized in one thing. These actions can be easily 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 fusing the described animation layers. For instance, there are three basic actions defined in Fig. 5 to create a «question about a big cat». Basic SLERP interpolation (Watt & Watt, 1992) is also used to concatenate signs smoothly in an utterance.

The realistic result of the movements is probably the most important elements to consider in the representation of sign language. The results obtained in ConSignos improve the results obtained in similar systems thanks to the use of the realistic rendering approach and the composition of individual actions. The keyframe 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 ConSignos can be easily used through distributed rendering approaches (Gonzalez-Morcillo, Weiss, Vallejo, Jimenez, & Castro-Schez, 2010).

6.2. Tools for increasing its adaptability

Social responses to virtual humans have been studied, using objective and subjective methods, in different contexts. The behavioral 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, Heloir, & Nguyen, 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 ConSignos the representation of the internal IK skeleton is shared among virtual characters using a XML specification. This file also specifies the relative size of the bones and the constraints required to generate realistic movements. Fig. 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 among avatars, the definition of each sign must be done only once. In previous developments of this representation module (Herrera, Gonzalez-Morcillo, Garcia, Mateos J. A., & I., 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 addking keyframes using the appropriate rotation value.

One of the main problems related to the creation of the signs is the time required to be modelled. In spite of the development of new techniques to ease the animation of virtual characters (such as inverse kinematics controls and key poses), the user may spend between 15 and 30 min to setup a new sign. It is important to recall that each sign must be done 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 can be considered as the main bottleneck in the project.

A sign editor module (Fig. 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 inversed kinematics facilities previously discussed. These configurations of the shape and orientation are defined over static poses which contains only the essential parameters that describe the action. For example, in Fig. 8, to describe the selected orientation of the hand, only the rotation of the bone hand.r and the location and rotation of

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

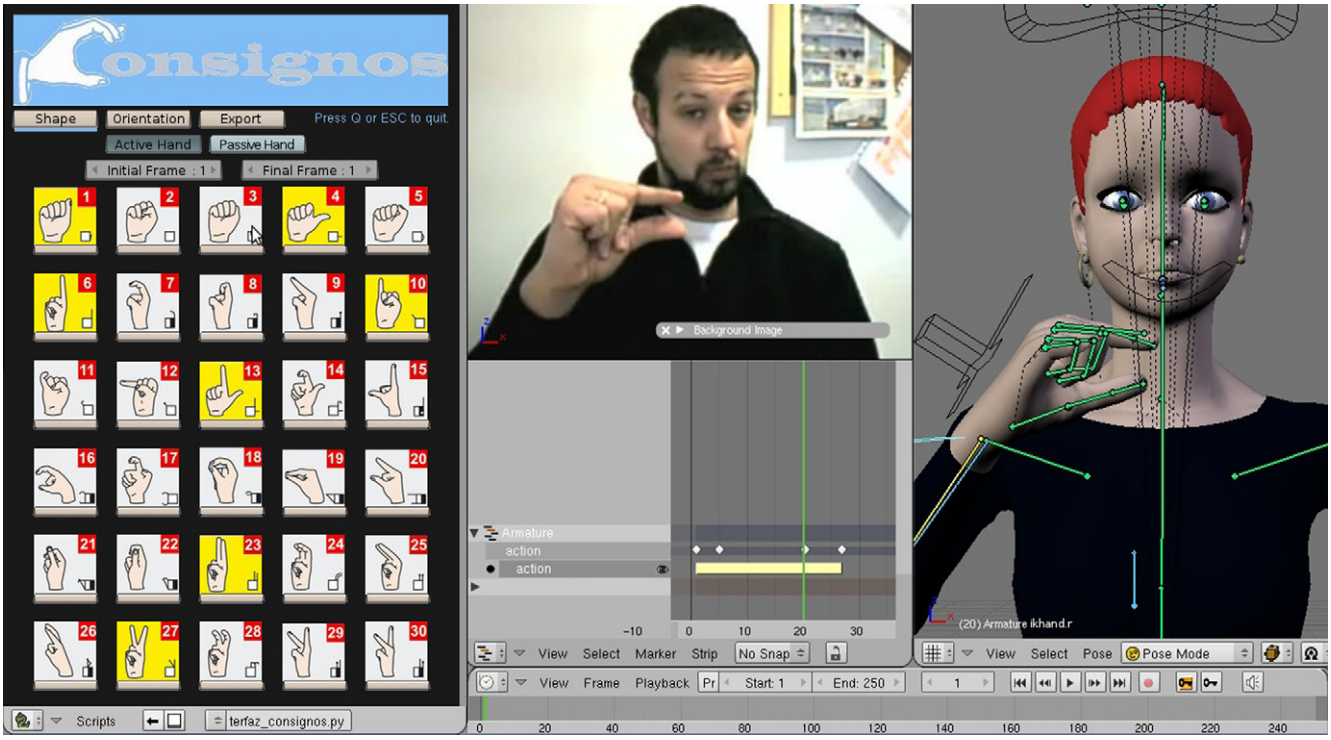

Fig. 7. Sign editor based on the use of predefined static poses for hand shape and orientations.

```
<?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>
```

Fig. 8. Static pose definition in XML of a right hand orientation setup.

ikhand.r are relevant. This information is stored in XML files. Fig. 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. Fig. 7 shows the first 30 configurations in the sign editor. For example, as shown in Fig. 7, shape 6 (background in yellow colour) defines one basic configuration (index finger stretched) and shapes from 7 to 8 are derived from it (different configuration with the same finger stretched). Thanks to the use of this sign editor, the time required to specify

Table 2
Analysis of effort reduction (in PM person x month) when adapting the system to a new domain thanks to the new tools and utilities developed in this work.

Effort in PM (person x month)		
Previous version	Current version	Task
2 PM	2 PM	Parallel corpus development
0.25 PM		Source language variants generation
2 PM		Models for translation
2 PM	0.5 PM	Sign design and development
6.25 PM	2.5 PM	Total

a new sign decreased the 90% with similar quality results. Some examples can be downloaded from http://www.esi.uclm.es/www/cglez/ConSignos/signos/.

7. Discussion and conclusions

Table 2 presents an analysis of the effort reduction (in PM person × months) when adapting the system to a new domain using the new tools and characteristics incorporated in the new version of a Spanish into LSE translation system.

As it is shown, the total effort has been reduced significantly (more than 50%) thanks to the elimination of the effort necessary to develop the translation rules manually (now, all the information for language translation is generated automatically), and the important reduction in the sign design and development process. The new version of the system not only increases its adaptability by reducing the required effort to be adapted to a new domain, but also, introducing new possibilities (like the acoustic adaptation module for the speech recognizer) that permit to maintain the system performance in different conditions.

As a conclusion, this paper has presented a new version of a speech into sign Language translation system with new tools and characteristics for increasing its adaptability to a new task or a new semantic domain. This paper presents new improvements in all the three main modules. The speech recogniser includes an acoustic adaptation module, for adapting the acoustic models to a new specific environment (indoor or outdoor scenarios), a new speaker, or a new Spanish accent. When generating automatically the vocabulary and a language model for the speech recogniser, a new module has been included for introducing source language variants, increasing the speech recogniser flexibility. The language translation module presents a new configuration compared to the previous version (San-Segundo et al., 2012) where all the translation strategies are data-oriented ones. With this new design, the required models are generated automatically from a parallel corpus. The statistical translation strategy incorporates a new pre-processing module (López-Ludeña et al., 2012) that permits to increase its performance, allowing replacing the rule-based translation strategy. The sign animation module includes a new version of the sign editor that incorporates new options (like predefined positions and orientations) for reducing significantly the sign specification time.

The whole system presents a SER lower than 10% and a BLEU higher than 90% for a specific domain application.

Acknowledgment

This work has been supported by Plan Avanza Exp N: TSI-020100-2010-489 and the European FEDER fund.

References

Adamo-Villani, N. (2008). 3D Rendering of American sign language finger spelling: A comparative study of two animation techniques. In *5th international conference on computer instructional technologies* (pp. 808–812). Italy.

Brown, P. F., Della Pietra, V. J., deSouza, P. V., Lai, J. C., Mercer, R. L., et al. (1990). Class based n gram models of natural language. *Computational Linguistics, 18*(4), 467–469.

Brown, R. D. (2000). Automated generalization of translation examples. In *Proceedings of the eighteenth international conference on computational linguistics (COLING-2000)* (pp. 125–131). Saarbrücken, Germany, August 2000.

Conroy, P. (2006). *Signing in and Signing out: The education and employment experiences of Deaf adults in Ireland*. Dublin: Irish Deaf Society. Research Report.

Cox, S. J., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Mand Tutt, & Abbott, S. (2002). TESSA, a system to aid communication with deaf people. In *ASSETS 2002* (pp. 205–212). Edinburgh, Scotland.

Dreuw, P., Neidle, C., Athitsos, V., Sclaroff, S., & Ney, H. (2008). Benchmark databases for video-based automatic sign language recognition. In *International conference on language resources and eval-uation (LREC)* (pp. 1115–1121). Marrakech, Morocco, May 2008.

Efthimiou, E., & Fotinea, E. (2008). GSLC: Creation and annotation of a Greek sign language corpus for HCI. *LREC, 2008*, 1–10.

Elliott, R., Glauert, J. R. W., Kennaway, J. R., Marshall, I., & Safar, E. (2008). Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society* (Vol. 6(4), pp. 375–391): Springer.

Ferreiros, J., San-Segundo, R., Fernández, F., D'Haro, L., Sama, V., Barra, R., & Mellén, P. (2005). New word-level and sentence-level confidence scoring using graph theory calculus and its evaluation on speech understanding. In *Interspeech 2005* (pp. 3377–3380). Lisboa, Portugal, September 2005.

Filhol, M. (2009). Zebedee: A lexical description model for sign language synthesis, In *Limsi internal report 2009–08*. Orsay.

Gauvain, J. L., & Lee, C. H. (1994). Maximum a-posteriori estimation for multivariate Gaussian mixture observations of Markov chains. *IEEE Transactions on SAP, 2*, 291–298.

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 is-sues and some preliminary results. In *4th workshop on the representation and processing of sign languages: Corpora and sign language technologies (CSLT 2010)* (pp. 98–102). Valletta, Malta, May 2010.

Gonzalez-Morcillo, C., Weiss, G., Vallejo, D., Jimenez, L., & Castro-Schez, J. J. (2010). A multiagent architecture for 3D rendering optimization. *Applied Artificial Intelligence, 24*(4), 313–349.

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: Corpora and sign language technologies (CSLT 2010)* (pp. 106–110). Valletta, Malta, May 2010.

Hermansky, H. (1990). Perceptual linear prediction (PLP) analysis for speech. *Journal of the Acoustical Society of America, 87*(4), 1738–1752.

Herrera, V., Gonzalez-Morcillo, C., Garcia, M. A., Mateos J. A., & Arriaga, I. (2009). Ganas: A flexible architecture for 3D sign language rendering. In *International conference interfaces human and computer interaction* (pp. 61–68). Portugal.

Huenerfauth, M., & Hanson, V. L. (2009). Sign language in the interface. Access for deaf signers. *The Universal Access Handbook*.

Jaballah, K., & Jemni, M. (2010). Toward automatic sign language recognition from Web3D based scenes. Lecture Notes in Computer Science, 6190, pp. 205–212: Springer.

Jankowski, C. R., Hoang-Doan Jr., & Lippmann, R. P. (1995). A comparison of signal processing front ends for automatic word recognition. *IEEE Transactions on Speech and Audio Processing, 3*(4), 286–293.

Karami, A., Zanj, B., & Sarkaleh, A. (2011). Persian sign language (PSL) recognition using wavelet transform and neural networks. *Expert Systems with Applications, 38*(3), 2661–2667.

Kipp, M., Heloir, A., & Nguyen, Q. (2011). *Sign language avatars: Animation and comprehensibility*. *Intelligent Virtual Agents*. Springer, pp. 113–126.

Koehn, P., Och, F. J., & Marcu, D. (2003). Statistical phrase-based translation. In *Human language technology conference 2003 (HLT-NAACL 2003)* (pp. 127–133). Edmonton, Canada.

Koehn, P. (2010). *Statistical machine translation*. phD. Cambridge University Press.

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

Lever, N. (2002). *Real time 3D character animation with visual C++*. Focal Press.

López-Ludeña, V., San-Segundo, R., Montero, J. M., Córdoba, R., Ferreiros, J., & Pardo, J. M. (2012). Automatic categorization for improving Spanish into Sign Language Translation System. *Computer Speech and Language, 26*(3), 149–167.

Marshall, I., & Sáfár, E. (2005). Grammar development for sign language Avatar-based synthesis. In *Proceedings HCII 2005, 11th International Conference on Human Computer Interaction (CD-ROM)* (pp. 1–10). Las Vegas, USA, July 2005.

Moreno, A. (1997). SpeechDat Spanish database for fixed telephone networks. Corpus design technical report, SpeechDat project LE2-4001.

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.

Morrissey, S., Way, A., Stein, D., Bungeroth, J., & Ney, H. (2007). Towards a Hybrid Data-Driven MT Sys-tem for sign languages. Machine translation summit (MT Summit), Copenhagen, Denmark, pp. 329–335.

Morrissey, S. (2008). *Data-Driven machine translation for sign languages*. Thesis. Dublin, Ireland: Dublin City University.

Munib, Q., Habeeb, M., Takruri, B., & Al-Malik, H. (2007). American sign language (ASL) recognition based on Hough transform and neural networks. *Expert Systems with Applications, 32*(1), 24–37.

Niewiadomski, R., Bevacqua, E., Mancini, M., & Pelachaud, C. (2009). Greta: An interactive expressive ECA system. In *8th international conference on autonomous agents and multiagent systems* (Vol. 2, pp. 1399–1400).

Och, J., & Ney, H. (2003). A systematic comparison of various alignment models. *Computational Linguistics, 29*(1), 19–51.

Ohali, Y. A. (2010). Identification of most desirable parameters in SIGN language tools: A comparative study. *Global Journal of Computer Science and Technology, 10*(6), 23–29.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. In *40th annual meeting of the association for computational linguistics (ACL)* (pp. 311–318). Philadelphia, PA.

Paulson, L. D. (2008). News briefs. *IEEE Computer Magazine, 41*(2), 23–25.

Pfau, R., & Quer, J. (2010). *Nonmanuals: Their grammatical and prosocid roles. Sign languages*. Cambridge University Press, pp. 381–402.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T., & Henning, J. (1989). *HamNoSys version 2.0: Hamburg notation system for sing language*. Signum.

San-Segundo, R., Barra, R., Córdoba, R., D'Haro, L. F., Fernández, F., Ferreiros, J., et al. (2008). Speech to sign language translation system for Spanish. *Speech Communication, 50*(2008), 1009–1020.

San-Segundo, R., Montero, J. M., Córdoba, R., Sama, V., Fernández, F., D'Haro, L. F., et al. (2012). Design, development and field evaluation of a Spanish into sign language translation system. *Pattern Analysis and Applications, 15*(2), 203–224.

San-Segundo, R., Pardo, J. M., Ferreiros, F., Sama, V., Barra-Chicote, R., Lucas, J. M., et al. (2010). Spoken Spanish generation from sign language. *Interacting with Computers, 22*(2), 123–139.

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.

Stein, D., Bungeroth, J., & Ney, H. (2006). Morpho-Syntax based statistical methods for sign language translation. In *11th annual conference of the Euro-pean association for machine translation* (pp. 223–231). Oslo, Norway, June 2006.

Stolcke, A. (2002). "SRILM – An extensible language modelling toolkit. In *Proceedings of International conference on spoken language processing* (Vol. 2, pp. 901–904). Denver, USA.

Sylvie, O., & Surendra, R. (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(6).

Thiebaux, M., Marsella, S., Marshall, A. N., & Kallman. M. (2008). Smartbody: Behavior realization for embodied conversational agents. In *7th international joint conference on autonomous agents and multiagent systems* (Vol. 1, pp. 151–158).

Vendrame, M., Tiotto, G. (2010). ATLAS Project: Fore-cast 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)* (pp. 239–243). Valletta, Malta, May 2010.

Watt, A., & Watt, M. (1992). *Advanced animation and rendering techniques*. UK: Pearson Education Harlow.

Wheatley, M., & Pabsch, A. (2010). Sign Language in Europe. In *4th workshop on the representation and processing of sign languages: Corpora and sign language technologies.* W. Stokoe, sign language structure: An outline of the visual communication systems of the American deaf, studies in linguistics, Buffalo University Paper 8, 1960. LREC, Malta 2010.

Wolfe, R., McDonald, J., Davidson, M., & Frank, C. (2007). Using an animation-based technology to support reading curricula for deaf elementary schoolchildren. In *22nd annual international technology & persons with disabilities conference*. Los Angeles, March 2007.