Interacting with Computers xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Interacting with Computers

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

Spoken Spanish generation from sign language

R. San-Segundo [a,*], J.M. Pardo [a], J. Ferreiros [a], V. Sama [a], R. Barra-Chicote [a], J.M. Lucas [a], D. Sánchez [b], A. García [b]

[a] Departamento de Ingeniería Electrónica, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain
[b] Fundación CNSE, Confederación Estatal de Personas Sordas, Spain

ARTICLE INFO

Article history:
Received 5 June 2009
Received in revised form 10 November 2009
Accepted 25 November 2009
Available online xxxx

Keywords:
Spanish sign language (LSE)
Speech generation from LSE
LSE corpus
Sign editor
LSE translation
Driver's license renewal

ABSTRACT

This paper describes the development of a Spoken Spanish generator from sign-writing. The sign language considered was the Spanish sign language (LSE: Lengua de Signos Española). This system consists of an advanced visual interface (where a deaf person can specify a sequence of signs in sign-writing), a language translator (for generating the sequence of words in Spanish), and finally, a text to speech converter. The visual interface allows a sign sequence to be defined using several sign-writing alternatives. The paper details the process for designing the visual interface proposing solutions for HCI-specific challenges when working with the Deaf (i.e. important difficulties in writing Spanish or limited sign coverage for describing abstract or conceptual ideas). Three strategies were developed and combined for language translation to implement the final version of the language translator module. The summative evaluation, carried out with Deaf from Madrid and Toledo, includes objective measurements from the system and subjective information from questionnaires. The paper also describes the first Spanish-LSE parallel corpus for language processing research focused on specific domains. This corpus includes more than 4000 Spanish sentences translated into LSE. These sentences focused on two restricted domains: the renewal of the identity document and driver's license. This corpus also contains all sign descriptions in several sign-writing specifications generated with a new version of the eSign Editor. This new version includes a grapheme to phoneme system for Spanish and a SEA-HamNoSys converter.

1. Introduction

Based on information from Spanish Statistics Institute[1] (INE) and the Ministry of Education[2] (MEC), around 47% of the Deaf, older than 10, do not have basic level studies or are illiterate. In real conditions, 92% of the Deaf have significant difficulties in understanding and expressing themselves in written Spanish. The main problems are related to verb conjugations, gender/number concordances and abstract concepts explanations. Because of this, only between 1% and 3% of the Deaf have a university degree. This percentage is very low compared to all the population in Spain.

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 like the Fundación CNSE, LSE is becoming not only the natural language for the Deaf to communicate with, but also a powerful instrument when communicating to people who can hear, or accessing information.

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

The paper is organized as follows: Section 2 reviews the state of art. Section 3 describes the first Spanish-LSE parallel corpus for language processing research. Section 4 presents the spoken Spanish generation system: module description and summative evaluation. The main conclusions and the future work are described in Sections 5 and 6.

* Corresponding author. Address: Grupo de Tecnología del Habla, Dpto. Ingeniería Electrónica, ETSI Telecomunicación, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain. Tel.: +34 915495700x4228; fax: +34 913367323.
E-mail address: lapiz@die.upm.es (R. San-Segundo).

[1] http://www.mec.es.
[2] http://www.let.kun.nl/sign-lang/echo/.

0953-5438/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.intcom.2009.11.011

Table 1
The main corpora in the international community.

References	Sign language	Description
http://www.let.kun.nl/sign-lang/echo/	Swedish sign language (SSL), British sign language (BSL) and the sign language of The Netherlands (SLN)	The European Cultural Heritage Online organisation (ECHO) corpora contain children's fables and poetry each signed by a single signer. This corpus covers a very wide language domain with a very large vocabulary which makes language processing and learning difficult
http://www.ru.nl/corpusngtuk/	Sign language of The Netherlands (SLN)	The Corpus NGT is an open access online corpus of movies with annotations of sign language of The Netherlands. Data were collected from native signers, preferably signers who have one or two deaf parents or even more deaf family members. The signers are all adults, their ages ranging from 17 to 84 years of age. The signers come from all over the country
Leeson et al. (2006)	Ireland sign language (ISL)	The Signs of Ireland corpus developed at the Centre for Deaf Studies, Dublin contains video data of approximately 40 Deaf ISL users collected over 3 years. Participants aged between 18 and 65 tell personal narrative, a children's story and sign elicited sentences. The corpus was hand annotated using ELAN software
Su et al. (2007)	Chinese and Taiwanese sign language	Chiu and Cheng performed Statistical Machine Translation (SMT) experiments on a corpus of about 2000 sentences for the language pair Chinese and Taiwanese sign language
Bungeroth et al. (2006)	German sign language (DGS)	This database contains a corpus of 2468 sentences in German and German sign language (DGS) for the domain of weather reports. It is particularly used for SMT and sign language recognition
Neidle et al. (2000)	American sign language (ASL)	The American sign language Linguistic Research group at Boston University created a set of videos in American sign language which is partly available on their website.[a] In 2007 this corpus has been increased: the new data release includes 15 short narratives (vols. 3–7) as well as over 200 additional elicited utterances (vol. 2)
Bungeroth et al. (2008)	English, German, Irish sign language, German sign language and South African sign language	The ATIS corpus is based on the domain of air travel information. It is available for five languages, English, German, Irish sign language, German sign language and South African sign language. The domain is restricted to flight information and booking services. From the original ATIS language corpus, 595 sentences were chosen for the sign language translation
Johnston (2008)	Australian sign language	This corpus has more than 300 h from 100 speakers in Australian sign language (it comes from British and New Zealand sign language)
Dreuw et al. (2008a)	American sign language (ASL)	The RWTH-BOSTON-400 Database contains 843 sentences (continuous sign language), with about 400 different signs from 5 speakers. These are in American sign language with English annotations. This corpus is mainly used for automatic sign language recognition
Schembri (2008)	British sign language (BSL)	The British sign language Corpus Project is a new 3-year project (2008–2010) that aims to create a machine-readable digital corpus of spontaneous and elicited British sign language (BSL) collected from deaf native signers and early learners across the United Kingdom. Although the recruitment of participants is being balanced for gender and age, it focus only on signers exposed to BSL before the age of 7 years, and adult deaf native signers will be disproportionately represented. Signers will also be filmed in 8 key regions across the United Kingdom, with a minimum of 30 participants from each region
Campr et al. (2008)	Czech sign language (CSL)	The UWB-07-SLR-P corpus contains video data of 4 signers recorded from 3 different perspectives for Czech sign language. Two of the perspectives contain whole body and provide 3D motion data, the third one is focused on signer's face and provide data for face expression and lip feature extraction. Each signer performed 378 signs with 5 repetitions. The corpus consists of several types of signs: numbers (35 signs), one and two-handed finger alphabet (64), town names (35) and other signs (244). In total, the corpus consists of 21,853 video files in total length of 11.1 h. The corpus is intended for training and testing of sign language recognition (SLR) systems
Efthimiou and Fotinea (2008)	Greek sign language (GSL)	This corpus has been developed at Institute for Language and Speech Processing (ILSP) and it contains parts of free signing narration, as well as a considerable amount of grouped signed phrases and sentence level utterances (18 h of signing data).

[a] http://www.bu.edu/asllrp/.

2. State of art

This section includes the state of art of the main subjects tackled in this paper: speech generation from sign language and sign language corpora.

2.1. Main corpora for sign language research

The research into sign language has been possible thanks to corpora generated by several groups. This section summarizes the main corpora in the international community. Some of them focus on linguistic aspects instead of language processing as it is needed in language translation. This description has been provided in Table 1.

For Lengua de Signos Española (LSE), there is a lot of content on videos but they are not annotated. Some examples are the efforts in Vigo University[3] and in the Instituto Cervantes.[4] The Fundación CNSE has developed the first normative dictionary for LSE (DILSE III) with more than 4000 signs, including video, description and definition.[5] Because of the lack of LSE, this paper presents the first Spanish-LSE parallel corpus for language processing research focused on specific domains. This corpus includes more than 4000 Spanish sentences translated into LSE. These sentences are focus on two restricted domains: the renewal of the identity document and driver's license. This paper also presents the design, for avatar

[3] http://webs.uvigo.es/lenguadesignos/sordos/lineas/index.htm.

[4] http://www.cervantesvirtual.com/seccion/signos/.

[5] http://www.fundacioncnse.org/tesorolse.

Fig. 1. Example of characteristics of a sign using HamNoSys.

representation, of all signs included in the database. A new version of the eSIGN editor has been used for sign design.

About the different ways of writing a sign (sign-writing), traditionally the sign has been written using words (in capital letters) in Spanish (or English in the case of BSL, British sign language) with a similar meaning to the sign meaning. They are called glosses (i.e. 'HOUSE' for the sign 'house'). In the last 20 years, several alternatives, based on specific characteristic of the signs, have appeared in the international community: HamNoSys (Prillwitz et al., 1989) (Fig. 1), Sistema de Escritura Alfabética (SEA) (Herrero, 2004) and sign writing (http://www.signwriting.orgl). But nowadays, neither of these options is well known by the Deaf in Spain, so the glosses seem to be the best option for LSE writing.

2.2. Generating spoken language from sign language

In order to permit a full dialogue (allowing communication in both directions) between a hearing person and a deaf person, it is necessary not only to translate speech into sign language (San-Segundo et al., 2008) but also to generate spoken language from sign language. In previous projects such as VANESSA[6] (Tryggvason, 2004), this problem has been solved by asking the Deaf to write down the sentence in English (or Spanish in our case) and then a text to speech (TTS) converter can generate the speech. But this is not a good solution because a very high percentage of the Deaf cannot write well in Spanish. Sign language is their first language, and their ability to write or read may be poor in many cases.

Because of this, a great deal of effort has been made in recognising sign language and translating it into spoken language by using a language translator and a TTS converter. Some efforts have been oriented towards recognising signs from the information captured with specialised hardware (Yao et al., 2006). However, this is an invasive approach which is sometimes rejected by the Deaf.

Others efforts try to recognize signs from video processing. The systems developed so far are very person or environment dependent (Vogler and Metaxas, 2001), or they focus on the recognition of isolated signs (Wang et al., 2006; von Agris et al., 2006) which can often be characterised just by the direction of their movement. In Sylvie and Surendra (2005), a review of research into sign language and gesture recognition is presented.

Dreuw is making a significant effort into recognizing continuous sign language from video processing (Dreuw, 2008; Dreuw et al., 2008a,b, 2009): the results obtained are very promising but this technology is not ready yet for developing real systems for the Deaf. The best experiment reported by Dreuw et al. (2008c) showed a sign recognition error rate of 17.5% using the RWTH-Boston-104 Database. In these experiments, the system tries to recognized sign sequences represented by three people continuously (without pauses between signs). The number of signs considered was 106 signs. Considering isolated signs, it is possible to reduce the recognition error to under 10% when considering vocabularies of a few hundred signs (Sylvie and Surendra, 2005). Once the sign language is recognized, the sign sequence is translated into a word sequence which is then passed to a TTS converter to generate the speech. Signs are represented by glosses, capitalised words with a semantic relationship to sign meaning.

Because sign recognition technology is not mature enough, this paper describes an advanced interface where a deaf person can specify a sign sequence (gloss sequence) with a set of visual tools. This sequence is translated into words that will be spoken. This solution allows the Deaf to specify a sentence in their first language (LSE) and avoid errors from sign recognition. As is shown in the summative evaluation, this interface has been well assessed by the users. This is the first system developed for Spanish and it complements the previous efforts into developing a Spanish into LSE translation system (San-Segundo et al., 2008).

3. Spanish-LSE parallel corpus

This section describes the first Spanish-LSE parallel corpus developed for language processing in two specific domains: the renewal of the identity document (ID) and driver's license (DL). This corpus has been obtained with the collaboration of Local Government Offices where these services are provided. Over several weeks, the most frequent explanations (from the government employees) and the most frequent questions (from the user) were taken down. In this period, more than 5000 sentences were noted and analysed.

Not all the sentences refer to ID or DL renewal (Government Offices provide more services), so sentences had to be selected manually. This was possible because every sentence was tagged with the information about the service being provided when it was collected. Finally, 1360 sentences were collected: 1023 pronounced by government employees and 337 by users. These sentences have been translated into LSE, both in text (sequence of glosses) and in video, and compiled in an excel file. This corpus was increased to 4080 by incorporating different variants for Spanish sentences (maintaining the LSE translation) (Fig. 2).

The Excel file contains eight different information fields: "ÍNDICE" (sentence index), "DOMINIO" (domain: ID or DL renewal), "VENTANILLA" (window: where the sentence was collected), "SERVICIO" (service provided when the sentence was collected), if the sentence was pronounced by the government employee or user (funcionario or usuario respectively), sentence in Spanish (CASTELLANO), sentence in LSE (sequence of glosses), and link to the video file with LSE representation (Fig. 3).

The main features of the corpus are summarised in Table 2. These features are divided depending on the domain (identity document or driver's licence renewal) and whether the sentence was spoken by the government employee or the user.

In the parallel corpus, all signs were written using glosses (capitalised words with a semantic relationship to sign language). In order to consider other sign-writing notations, a database with 715 signs (including all signs in the parallel corpus) was generated. This database includes (Fig. 5) sign descriptions in:

- Glosses (GLOSA column): word in capital letters. For example, ABAJO (DOWN).
- Sistema de Escritura Alfabética (SEA). ABAJO sign is represented with the following characteristics: ole (hand shape) mua (hand orientation) wu (top-down movement).
- HamNoSys. The HamNoSys characteristics for ABAJO sign are (Fig. 4).

6 http://www.visicast.cmp.uea.ac.uk/eSIGN/Vanessa.htm.

Please cite this article in press as: San-Segundo, R., et al. Spoken Spanish generation from sign language. Interact. Comput. (2009), doi:10.1016/j.intcom.2009.11.011

INDICE	DOMINIO	VENTANILLA	SERVICIO	TIPO	CASTELLANO	LSE	VIDEO
1	DL	CAJA	Decir la cantidad	Funcionario	ahora tiene que ir a la ventanilla de conductores	AHORA TU VENTANILLA ESPECÍFICO PERSONA CONDUCTOR IR-ALLÍ	1.wmv
2	DL	CAJA	Decir la cantidad	Funcionario	ahora vaya a la ventanilla de conductores	AHORA TU VENTANILLA ESPECÍFICO PERSONA CONDUCTOR IR-ALLÍ	2.wmv
3	DL	CAJA	Decir la cantidad	Funcionario	catorce con veinte	CATORCE COMA VEINTE EUROS	3.wmv
4	DL	CAJA	Decir la cantidad	Funcionario	catorce euros con veinte céntimos	CATORCE COMA VEINTE EUROS	4.wmv
5	DL	CAJA	Decir la cantidad	Funcionario	ocho euros	OCHO EUROS	5.wmv
6	DL	CAJA	Decir la cantidad	Funcionario	si usted tiene más de setenta años no tiene que	EJEMPLO TU EDAD SETENTA MÁS CARNET CONDUCIR RENOVAR PAGA	6.wmv
7	DL	CAJA	Decir la cantidad	Funcionario	son catorce con veinte	CATORCE COMA VEINTE EUROS	7.wmv
8	DL	CAJA	Decir la cantidad	Funcionario	son catorce euros con veinte céntimos	CATORCE COMA VEINTE EUROS	8.wmv
9	DL	CAJA	Decir la cantidad	Funcionario	son ocho euros	OCHO EUROS	9.wmv
10	DL	CAJA	Decir la cantidad	Funcionario	vaya directamente a conductores	TU VENTANILLA ESPECÍFICO PERSONA CONDUCTOR IR-ALLÍ DIRECTO	10.wmv
11	DL	CAJA	Decir la cantidad	Funcionario	vaya directamente a ventanilla de conductores	TU VENTANILLA ESPECÍFICO PERSONA CONDUCTOR IR-ALLÍ DIRECTO	11.wmv
12	DL	CAJA	Decir la cantidad	Funcionario	ve directamente a conductores	TU VENTANILLA ESPECÍFICO PERSONA CONDUCTOR IR-ALLÍ DIRECTO	12.wmv
13	DL	CAJA	Decir la cantidad	Funcionario	ve directamente a la ventanilla de conductores	TU VENTANILLA ESPECÍFICO PERSONA CONDUCTOR IR-ALLÍ DIRECTO	13.wmv
14	DL	CAJA	Horarios	Funcionario	el horario de la ventanilla de caja es de nueve a	VENTANILLA ESPECÍFICO DINERO PAGAR HORA HORARIO ABRIR NUEV	14.wmv
15	DL	CAJA	Horarios	Funcionario	el resto de las ventanillas cierran a las dos	VENTANILLA LOS-DEMÁS CERRAR HORA DOS	15.wmv
16	DL	CAJA	Ir después	Funcionario	por favor vaya a la ventanilla de conductores	TU VENTANILLA ESPECÍFICO PERSONA CONDUCTOR IR-ALLÍ	16.wmv
17	DL	CAJA	Ir después	Funcionario	por favor ve a la ventanilla de conductores	TU VENTANILLA ESPECÍFICO PERSONA CONDUCTOR IR-ALLÍ	17.wmv

Fig. 2. Example of database content.

Fig. 3. Examples of videos recorded by the Fundación CNSE.

Table 2
Main statistics of the corpus generated.

	ID		DL	
	Spanish	LSE	Spanish	LSE
Government employee				
Sentence pairs	1425		1641	
Different sentences	1236	389	1413	199
Running words	8490	6282	17,113	12,741
Vocabulary	652	364	527	237
User				
Sentence pairs	531		483	
Different sentences	458	139	389	93
Running words	2768	1950	3130	2283
Vocabulary	422	165	294	133

SIGML (Zwiterslood et al., 2004) (the SIGML column includes a link to a text file with the SIGML description necessary for representing the sign with the eSIGN avatar[7]).

This database includes signs for all of the letters (necessary for word spelling), numbers from 0 to 100, numbers for hour specification, months, week days, and all signs in the parallel corpus. The sign database has been generated using a new version of the eSIGN editor (Hanke and Popescu, 2003). The eSIGN Editor was developed in the VISICAST and eSIGN European Projects (essential sign language information on Government networks). In this work, this editor has been adapted to LSE. The new version incorporates the same functionality for defining manual movements using HamNoSys and non-manual aspects such as movements of lips, head, etc. (Fig. 6). This new editor has three windows. In the main window, the eSign avatar shows the sign that is being designed at this moment (using a SEA or a HamNoSys specification). The second window allows HamNoSys characters to be introduced, and the last one permits non-manual gestures to be added. The SEA characters can be introduced using the PC–keyboard together with auxiliary buttons.

This new version incorporates two new features: a Spanish grapheme to phoneme tool and a SEA-HamNoSys converter. The Spanish grapheme to phoneme tool is a rule-based converter that generates a sequence of phonemes which are represented using speech assessment method phonetic alphabet (SAMPA) (Wells et al., 1997), given a Spanish sentence.[8] This sequence is necessary to make the avatar move the lips according to this pronunciation. This grapheme to phoneme is the same as those used in the automatic speech recognizer but adapted to generate SAMPA phonemes (required by the avatar specification).

The second feature is the SEA-HamNoSys converter. Sistema de Escritura Alfabética (SEA) (Herrero, 2004) is an alphabet (based on ASCII characters) for sign language. Like HamNoSys, SEA allows a sequence of characters to be specified that describe aspects such as hand-shape, hand position, location and movements. The reason for developing this converter is that the first normative dictionary for LSE (developed at Fundación CNSE[9]) has SEA descriptions for more than 4000 signs, but the eSign avatar needs HamNoSys descriptions for sign representation. It is important to comment that both notations (SEA and HamNoSys) only specify the manual part of a sign (hand movements) but they do not define other non-manual aspects such as face, head or body movements.

The SEA-HamNoSys converter (see Fig. 7) has been implemented in three steps: SEA characteristic detection, SEA-HamNoSys conversion for individual characteristics, and the generation of HamNoSys sign descriptions. These steps are repeated for all syllables that make up the sign, if there is more than one.

For SEA characteristic detection, seven different types of sign have been defined: five types for two-hand signs (differentiating whether the movement is symmetric or if it is done by the dominant or non-dominant hand) and two types for one-hand signs (done by the dominant or non-dominant hand). Depending on the type of sign, SEA sign characteristics are specified in a different order. Given a sign, it is necessary to detect the sign type in order to get its characteristics from the SEA description. Some of the sign characteristics can be omitted in SEA specifications. In these cases, a default value has to be considered. The main problem appears when the default value assigned to a specific characteristic depends on the value of other characteristics: for example, the default value for hand orientation depends on the current hand position. In order to assign default values, several programming rules were incorporated.

When all the SEA characteristics have been extracted, the next step is to convert them into HamNoSys. This conversion is done by using conversion tables. The system incorporates different tables for converting sign characteristics related to hand position, hand shape, hand orientation, movements, etc. As in the previous step, the main problem appears in those cases where a characteristic conversion depends on the value of other characteristics. In

7 http://www.sign-lang.uni-hamburg.de/esign/.

8 http://www.phon.ucl.ac.uk/home/sampa/spanish.htm.

9 http://www.signwriting.org/.

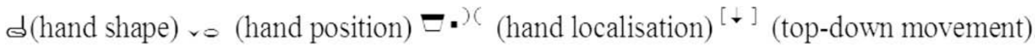

Fig. 4. HamNosys structure for ABAJO sign.

GLOSA	SEA	HAMNOSYS	SIGML
ABAJO	olemuawu	[illegible]	SIGML\ABAJO.txt
ACOMPAÑAR-A_MI	saca íájwe-ye	[illegible]	SIGML\ACOMPAÑAR-A_MI.txt
ACTUAL	s omèawud	[illegible]	SIGML\ACTUAL.txt
ADIÓS	omaudahb	[illegible]	SIGML\ADIÓS.txt
ADJUNTAR	sòaméha òamèug	[illegible]	SIGML\ADJUNTAR.txt
AHÍ	elewe	[illegible]	SIGML\AHÍ.txt

Fig. 5. Partial view of the sign database generated.

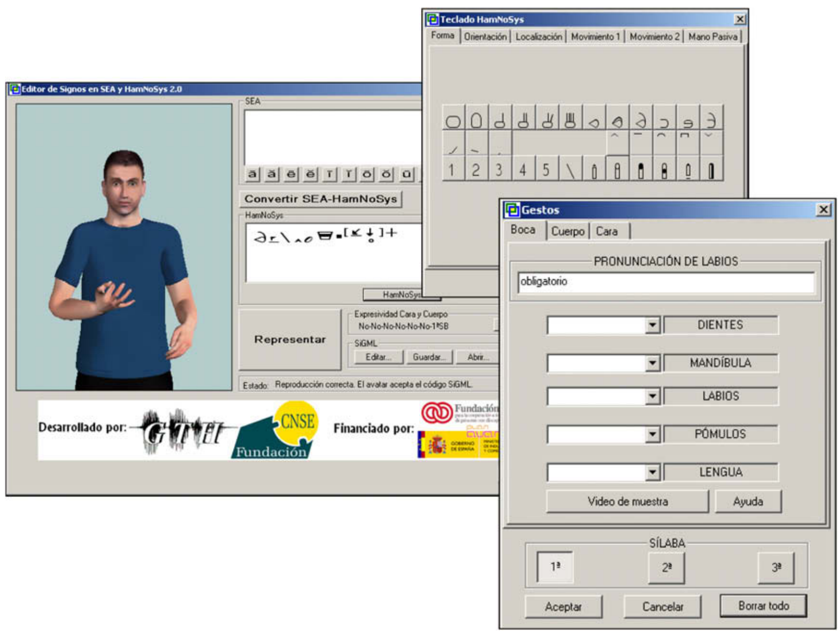

Fig. 6. New version of the visual sign editor. With the upper window, the user can specify the HamNoSys characteristics. With the "gestos" (gestures) window, the user can specify movements of mouth, body and face. In the case of the mouth it is possible to define de lip pronunciation, teeth, jaw, lips, cheek and tongue position.

this case, a new set of programming rules was defined for solving all these cases.

Finally, the SEA-HamNoSys converter generates a HamNoSys sign specification using the translated characteristics. For this generation, the system has defined seven different frames (corresponding to the seven types of sign defined in the first step). Depending on the sign type, a frame is selected and filled in for generating the HamNoSys specification.

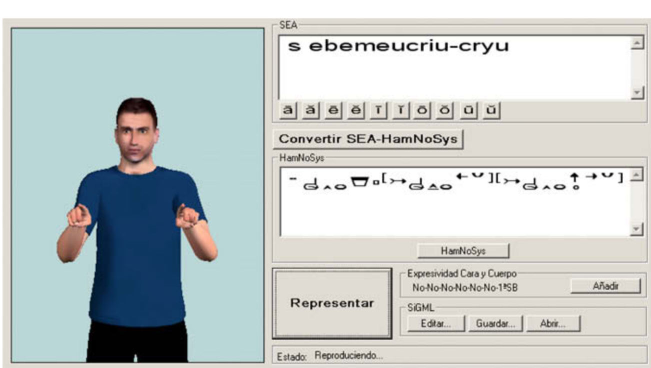

Fig. 7. SEA and HamNoSys representations of the sign MÚSICA (MUSIC).

The SEA-HamNoSys converter has been evaluated with 100 signs selected for including all the main SEA characteristics. The results have shown that for 75% of the signs, their HamNoSys specifications are very good. For the rest, it is necessary to make some modifications, but the HamNoSys structure generated is useful and syntactically correct. These results are due to the fact that these two sign-writing notations have different specification levels. SEA presents a higher level because it has been designed to be easy to learn. On the other hand, HamNoSys allows a very detailed level of sign design. Because of this, when converting from SEA to HamNoSys, it is sometimes necessary to incorporate additional information by making some assumptions that are not always correct.

For designing a sign, it is necessary to specify hand movements (manual part) and other gestures including face, head and body movements (non-manual part). For designing the manual part, two processes have been followed: if the sign was included in the normative dictionary from Fundación CNSE, its SEA description has been automatically converted into HamNoSys (and lightly modified if necessary). On the other hand, if the sign was not in the dictionary, the HamNoSys sign specification had to be generated from scratch, using the videos recorded by the Fundación CNSE as the reference. Most of the signs (around 70%) were included in the dictionary so the SEA-HamNoSys conversion tool

Please cite this article in press as: San-Segundo, R., et al. Spoken Spanish generation from sign language. Interact. Comput. (2009), doi:10.1016/j.intcom.2009.11.011

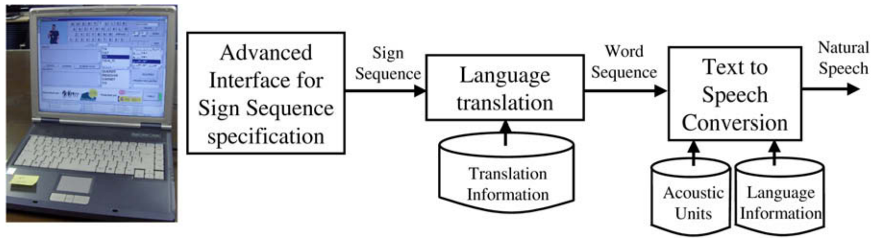

Fig. 8. Diagram module for the Spoken Spanish generation system.

has been very useful: the design time was reduced significantly, by approximately 50%. For the non-manual part of the sign, the design was always made from scratch, using the tools provided in the visual editor.

4. Spoken Spanish generation system

The spoken Spanish generation system has been developed based on three module architecture (Fig. 8). The main modules are as follows:

- The first step is to specify a sign sequence using an advanced visual interface. This interface includes several tools for sign specification: avatar for sign representation (to verify that sign corresponds to the gloss), prediction mechanisms, calendar and clock for date or time definitions, etc.
- The natural language translation module converts a sign sequence into a word sequence. For this module, three different strategies have been implemented and combined. The first one consists of an example-based strategy: the translation process is carried out based on the similarity between the sentence to translate and the items of a parallel corpus of translated examples. Secondly, a rule-based translation strategy, where a set of translation rules (defined by an expert) guides the translation process. The last one is based on a statistical translation approach where parallel corpora are used for training language and translation models.
- Finally, the word sequence is converted into spoken Spanish by using a commercial text to speech converter. In this project the Loquendo system was used.[10]

4.1. Design of the advanced visual interface

This section describes the process carried out in designing the advanced visual interface. During this process several analyses and tests were done involving the final users.

The main objective was to develop a visual interface for helping the Deaf to specify a sign sequence (a sentence in LSE). This design process has to be understood by considering its special context. This context has conditioned the design process and the alternatives considered. This context can be described in the following aspects:

- First of all, as was described in the introduction, a large percentage of the Deaf have significant difficulties in writing Spanish. This aspect conditions the interface to have an important visual component: the interaction is based on the PC–mouse (the user does not need to type).
- Spanish sign language (LSE: Lengua de Signos Española) has been considered one of the official languages in Spain recently since 2007. LSE is very young and the degree of standardization of this language is very low. This aspect can be observed in the different ways of representing the same sign (with the same meaning), and the different ways of writing a sign (sign-writing). The glosses seem to be the best option for LSE writing.
- Fundación CNSE (foundation of Spanish Deaf people) has started a significant number of activities for LSE normalization: defining standards, visual applications, LSE contents, etc. Perhaps the most important one is the generation of the first normative dictionary including more than 4000 signs (DILSE III: See footnote 5) with the sign description, meaning, video and associated gloss. In order to search for a sign, the dictionary offers two alternatives: one based on glosses (alphabetically) and the other based on several sign characteristics (hand shape, hand position, location, movements, etc.) similar to those aspects specified in the HamNoSys standard.
- Finally, LSE, like many other sign languages, presents a better sign coverage for describing tangible things, objects, etc., compared to abstract or conceptual ones. Most of the signs in the dictionary refer to tangible things. Because of this, it is more difficult to explain complex abstract or conceptual ideas to the Deaf. In most of the cases, the best way of explaining these ideas is describing a real example and defining analogies. This aspect will have an influence on the interface design, but also in the questionnaire design presented in the evaluation section, as will be shown.

The interface design process was carried out in three steps:

1. First, a group of experts had a brainstorming session to describe the necessary tools and different alternatives for designing the interface of each tool. This group was made up of two experts in LSE (Deaf and good representatives of potential users), a linguist (an expert in Spanish who also knows LSE) and a software developer (who does not know LSE). In the meeting, there were two interpreters for translating Spanish into LSE and vice versa.
2. Secondly, all the alternatives concerning the interface were implemented and evaluated with five final users (Deaf): formative evaluation. Every user tested all the design alternatives to generate 15 sentences in LSE using the interface. These 15 sentences were randomly selected from the database described in Section 3.
3. Finally, there was a meeting including the group of experts and the five users (including the two interpreters). During this meeting, the users were asked about their preferences, elaborating a detailed analysis and the final proposal for the interface.

4.1.1. Alternative designs

When designing a visual interface for specifying a sign sequence, the most important tool is the sign search. In this applica-

[10] http://www.loquendo.com/en/.

Please cite this article in press as: San-Segundo, R., et al. Spoken Spanish generation from sign language. Interact. Comput. (2009), doi:10.1016/j.intcom.2009.11.011

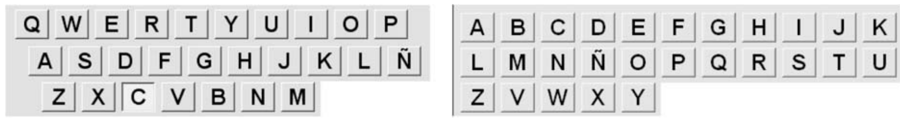

Fig. 9. Different distribution of letter keys for the sign search tool based on their glosses.

tion, there are more than 700 signs, so it is necessary to design an efficient mechanism for selecting a sign with just few clicks. In this meeting, most of the time was dedicated to discussing the different alternatives for the sign search.

The two first alternatives proposed were the same as those used in the DILSE III dictionary: glosses and specific characteristics of the sign.

For the first alternative, using glosses for the sign search, there were two proposals on the interface design (Fig. 9). In the first one, the keys were distributed like a traditional PC–keyboard, while in the second alternative, the keys were arranged alphabetically. The main idea has been to use a touch screen (although the mouse can be also used in all cases). The size of the button has been designed taking into account the possibility of using a touch screen.

For the second alternative, sign characteristic specification, the interface incorporated a submenu with seven windows where the user can specify sign characteristics related to the hand shape, position, location and movement. For characteristic specification, the HamNoSys standard was used, because it is the most widespread in the international community.

Additional to these two proposals, there were another two alternatives. The first one was to use videos (of one person signing) for every sign. The user has to play the videos to see which sign to select. The second proposal was to use images for representing the signs. In this case, the group of experts considered two possibilities: considering drawings or real pictures for representing the signs (Fig. 10). During the discussion, the real pictures were preferred because drawings can produce a feeling of children's interface, but both possibilities were tested.

The use of videos or pictures poses the problem that it is not possible to present many signs at the same time, and it is necessary to organize them in submenus. These submenus must have a hierarchical structure with several levels. Defining this organization has been one of the most complicated problems. In this case, based on the specific domain considered in this application, a syntactic–semantic organization was defined considering several submenus: objects like ("CAR"), documents ("APPLICATION FORM"), actions ("TO CARRY"), places ("DRIVER WINDOWS"), pronouns ("THIS"), adverbs ("MUCH"), persons ("OFFICER"), and others like numbers, hours, days, months, weekdays.

One aspect reported by one of the experts, and confirmed during the user tests was that this syntactic–semantic organization is not easy to explain to the Deaf for two reasons: the Deaf do not know the syntactic elements of Spanish very well and it is more difficult to express abstract ideas using LSE.

For the interface test, instead of generating videos, pictures and drawings for all signs in this domain (more than 700), only 50 videos, drawings and pictures were developed, including those necessary for composing the 15 sentences in the test.

Finally, the group of experts proposed other tools to be included in the interface:

- A calendar to specify a date. The signs corresponding to this date are automatically added to the sign sentence.
- A clock to specify an hour. The signs for specifying this hour are automatically added to the sign sentence.
- A list with the most used sign sentences: greetings, sentence to ask for repetition, etc.

Fig. 10. Example of drawing and picture of a car.

4.1.2. Formative evaluation: user tests and analysis of alternatives

All the alternatives concerning the interface were implemented and evaluated with the five final users. Table 3 presents the main characteristics of the five users involved in the formative evaluation. For 1 week, every user tested all the design alternatives to specify 15 sentences in LSE using the interface. These 15 sentences were randomly selected from the database.

After the user tests, the different alternatives were analysed by both the group of experts and the five users. Table 4 presents the analysis of the different strategies for designing the interface of the sign search tool.

As regards the calendar and the clock, the users proposed not to put the controls in the main interface because they were not used frequently: rather using submenus. This comment depends a lot on the 15 sentences selected for the test. In any case, it is true that in this domain there are not so many sentences including a date or a time.

Finally, the users proposed very interesting changes and new tools to be included in the interface:

- All the users proposed to increase the size of the letters in the interface.
- One of the users proposed a new tool for helping the user to specify the sign sequence. This new tool consists of proposing the three or four most probable next signs, given the last sign. For calculating the most probable signs, the interface can use the sign sequences from the database (Section 3). As will be shown in the evaluation, this tool has been very useful.
- Another tool proposed was to include a list of the last five previously used signs. Generally, the last used signs have a greater probability of being used again. With this list, it is not necessary to search for them again. As will be shown, this tool has not been as useful as the previous one.

4.1.3. Proposal for the visual interface

After the analysis of the different alternatives, the group of experts proposed a visual interface with the following characteristics, which justifies the proposal based on the analysis presented previously:

- For the sign search tool, the group of experts defined a mixed strategy based on glosses and videos. The sign search is performed over the glosses, specifying the first letter of the gloss. Then, all the glosses beginning with this letter are presented

Table 3
Characteristics of the five users involved in the formative evaluation.

User	Age	Literacy	Education level	Experience with computers
1	19	High	Bachelor	High
2	20	Medium	Basic	Low
3	37	High	Bachelor	High
4	42	Medium	Basic	Medium
5	45	Low	Basic	Low

in a list. Before adding a sign to the sign sequence, it is possible to see a video with the sign representation. Instead of recording one person signing all the signs, the decision was to include avatar animations. These animations have less quality than the videos, but the sign is perfectly identifiable and this solution increases the system scalability. Since the users did not report any differences between the two keyboard layouts, the standard PC layout was chosen, because this is familiar from computer keyboards.

- After a long discussion, the group of experts decided to complement the sign search by allowing the possibility of searching for a sign using HamNoSys. The main reason to support this decision was that in the normative dictionary there is a similar possibility for sign searching. As will be shown, it was not a good decision, because this tool was not used and it increased the interface complexity.
- The sign prediction characteristic was incorporated with a very high consensus with both users and experts. This tool has been one of the most used.
- The final proposal also included the list of the last five signs introduced. As will be shown, this list has not been very useful.
- In this domain, there are not so many sentences including a date or a time, so the decision was to include a calendar and a clock in submenus, not in the main window.
- The group of experts also proposed to include a list with the most frequent sign sentences (in a submenu): greetings, sentence to ask for repetition, etc.
- After this analysis, and thinking about the possibility of increasing the interface flexibility, a new tool was proposed for inclusion in the interface. This new tool consists of the possibility of specifying any gloss by spelling: by conforming a new gloss (not seen by the system before) letter by letter. This possibility is very interesting when introducing proper names. For this new tool, the buttons with the letters were reused: it was only necessary to include a new button to change the state: searching (default) or spelling.
- In order to facilitate the sign sequence specification, three buttons were proposed to be included: the first one to delete the last sign introduced, another button to delete the whole sequence, and finally, the third button for representing (using the avatar) the whole sign sequence (specified in glosses). Using the avatar for representing any sign sequence (without having it previously recorded) gives to the system a significant flexibility. The avatar concatenates the sign representations with smooth transitions between consecutive signs.
- Every time there is a change in the sign sequence, the translation from LSE to Spanish should be carried out automatically and the result presented in the word window. It is not necessary to have a button to carry out the translation. On the other hand, a button is necessary to speak the Spanish sentence (translation output).
- Finally, for political reasons, the interface had to include the sponsor logos.

4.2. Advanced visual interface for sign sequence specification

This section describes the main characteristics of the advanced visual interface developed in this work. This is a visual interface

Table 4
Analysis of alternatives for designing the interface of the sign search tool.

Alternative	Pros	Cons	Additional comments
Glosses	Same strategy as used in the normative dictionary DILSE III Easy to organize the search	The user needs to know words in glosses: words in Spanish The user must be familiar with the alphabetical order	There were no differences between the distribution of the two buttons (Fig. 9)
HamNoSys	Same strategy as used in the normative dictionary DILSE III Easy to organize the search	HamNoSys is not well known by the Spanish Deaf	None of the users knew HamNoSys
Videos	Videos is the natural way of representing signs of the LSE	It is very slow: it is necessary to play all signs in the submenu before selecting one The submenu structure is very difficult to design and very difficult to explain to deaf users There is ambiguity in the sign classification: one sign can be included in several submenus (i.e. objects and documents)	Two users reported the idea that when the user does not find a sign, there is a doubt whether the sign exists or whether it has been classified in another submenu
Pictures	Visual images help to identify a sign rapidly	Drawings can produce the feeling that the interface has been designed for children The interface can be associated to mentally disabled people or people with learning disabilities The submenu structure is very difficult to design and very difficult to explain to deaf users There is ambiguity in the sign classification: one sign can be included in several submenus (i.e. objects and documents)	All the users reported the idea that drawings are for children Two users were against the idea of representing signs with pictures or drawings. They associated the interface to children, mentally disabled people or people with learning disabilities. They demand a standard for sign-writing like other languages. They do not like the idea of writing a sentence in LSE (their mother language) by concatenating pictures

where the Deaf can build a sign sentence that it will be translated into Spanish and spoken to a hearing person. The sign sequence is specified in glosses but signs can be searched for using specific sign characteristics in HamNoSys notation. Fig. 11 shows the whole interface.

4.2.1. Gloss-based tools

The main utility consists of selecting a letter (clicking on one letter button, i.e. letter C in Fig. 11) and a list of glosses, beginning with this letter, is displayed in alphabetical order (these buttons have been included by considering the idea of using a touch screen to use the system, instead of the computer keyboard). If a gloss from the list is selected (i.e. CARNET in Fig. 11), the avatar (in the top-left corner) represents it to verify the wished sign corresponding to the gloss. In order to add this gloss to the sign sequence (the SIGNOS window, under the avatar in Fig. 11), it is necessary to click twice. On the right of the gloss list, there is a list of the corresponding HamNoSys notations.

At any moment, it is possible to carry out necessary actions: to represent the current sign sequence (SIGNAR button), to delete the last gloss introduced (BORRAR button) or to delete the whole sequence (BORRAR TODO button). Every time the sign sequence is modified, the language translation module is executed and the resulting word sequence is presented in the PALABRAS (words) window (under BORRAR buttons). The HABLA (speak) button executes the TTS converter over the word sequence specified in the PALABRAS (words) window. When the system is speaking, this button is disabled to avoid being used again.

By pressing the DELETREO (spelling) button, the system gets into the spelling state. In this state, the letter buttons have a different behaviour: they are used to introduce a gloss (in the SIGNOS window) letter by letter. In the example presented in Fig. 12, the gloss corresponding to a new proper name is included. When the avatar has to represent a new gloss, it checks whether there is a text file (in the gloss directory) with the sign description in SIGML corresponding to this gloss. If there are not any, the system signs letter by letter.

In order to generate several glosses by spelling, it is necessary to press the ESPACIO (space) button to separate consecutive glosses. In a similar way, it is also possible to specify a number using number buttons and the point button (for decimals).

A very useful tool incorporated into the interface allows proposing following glosses/signs given a previous sequence of glosses. When there is a partial gloss sequence specified and the user

Fig. 12. Example of introducing a new gloss by using the DELETREO (SPELLING) option: proper name MADRID.

moves the mouse over the SIGNOS windows, the system displays a popup menu proposing several candidates for the next gloss/sign (Fig. 13). These candidates have been selected based on a sign language model trained from gloss sequences (LSE sentences) in the corpus. The system proposes the four best signs: with highest probability of being selected, given the partial sequence already introduced. If the user clicks on one of these options, the gloss is incorporated into the sequence. This tool has been very useful as will be shown in the summative evaluation, Section 4.6.

4.2.2. HamNoSys-based search

As was commented on before, on the right of the gloss list, there is a list of the corresponding HamNoSys notations. The interface also allows signs to be searched for by specifying between 1 and 3 different HamNoSys characteristics. When these characteristics are specified, in the HamNoSys notation list, the sign notations including all these characteristics are presented. The gloss list (on the left) is updated corresponding to the HamNoSys notation list. In Fig. 14 an example of searching is presented.

The HamNoSys characteristics are selected by using a window with seven different tabs: similar to those used in the sign editor (see Fig. 6).

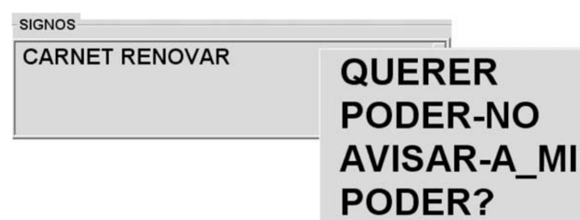

Fig. 13. Example of sign prediction. After including the signs CARNET (LICENCE) and RENOVAR (TO_RENEW), the most probable next signs are QUERER (TO_WANT), PODER-NO (CAN-NOT), AVISAR-A_MI (TELL_ME), PODER? (CAN?).

Fig. 11. Advanced visual interface for sign sequence specification. The sign sequence CARNET RENOVAR QUERER (LICENCE TO_RENEW TO_WANT) is translated into "me gustaría renovar el carné (I'd like to renew the driving licence).

Fig. 14. Example of searching using HamNoSys characteristics. The user has specified two HamNoSys characteristics. From the list of signs containing these characteristics, the user has selected the sign REGISTRO (REGISTRATION).

Fig. 15. Example of date and time introduction. In this example the user has selected the date April, 16th 2009 and the time 12:00.

4.2.3. General functions

Finally, several general functions are described in this section: date and time tools, list of the last glosses introduced, and frequent gloss sequences. In order to facilitate a date or time introduction, the interface offers the possibility of specifying the date in a calendar and the time on a digital clock (Fig. 15). When the date or the time has been specified, it is necessary to push the FECHA (date) or/and HORA (time) buttons to incorporate the corresponding glosses into the gloss sequence.

Another utility is the list of last five glosses used. Under the gloss list, the interface presents the last used glosses in order to facilitate the user selecting one of them. This utility has not been used very frequently in the summative evaluation.

Finally, the visual interface incorporates a list of the most frequent sign sequences (greetings, courtesy formulas, etc.). When one of these sequences is selected the whole sign sequence is replaced by the selected one, and the SIGNOS and PALABRAS windows are updated (Fig. 16).

This list with the most frequent sign sentences can be modified easily from a text file.

4.3. Language translation

The natural language translation module converts the sign sequence into a word sequence that will be spoken by the TTS converter. For this module, three different strategies have been implemented and combined in order to generate the final version of the translation module: example-based, rule-based and statistical translation.

For evaluating the rule-based and the statistical strategies, the following evaluation measurements have been considered: Word Error Rate (WER), Position Independent WER (PER), BiLingual Evaluation Understudy (BLEU Papineni et al., 2002), and NIST.[11] As is shown in Table 5, the rule-based strategy has provided better results in this task because it is a restricted domain and it has been possible to develop a complete set of rules with a reasonable effort. All of the sentences in the corpus have been checked manually and in all cases the word sequence was a correct sentence with the correct meaning. The amount of data for training the statistical models is very small so the models are not trained properly. In these situations, the rules defined by an expert introduce knowledge not seen in the data making the system more robust against new sentences.

4.3.1. Combining translation strategies

The natural language translation module implemented combines the three translation strategies that they will be described in next sections. This combination is set out in Fig. 17.

The translation module is divided into two main steps. At the first step, an example-based strategy is used for translating the gloss sequence. If the distance with the closest example is lower than a threshold (distance threshold), the translation output is the same as the example. But if the distance is higher, a background module is used to translate the gloss sequence. The distance threshold (DT) was fixed to 30% (one difference is permitted in a 4-sign sentence).

For the background module, a combination of rule-based and statistical translators has been used. Considering the results in Table 5, the rule-based strategy is the best alternative. In any event, the statistical approach was also incorporated as a good alternative during system development. During rule development, a statistical translator was incorporated in order to have a background module with a reasonable performance: a statistical translator can be generated in 2 days while the rule development takes several weeks. The relationship between these two modules is based on the ratio between the number of words generated after the translation process and the number of glosses in the input sequence. If this ratio is higher than a threshold the output is the word sequence proposed by the rule-based translator. Otherwise, the statistical approach is carried out. By analysing the parallel corpus, the ratio between number of words and number of glosses is 1.35. When the number of words generated by the rule-based approach is very low, it means that specific rules for dealing with this type of example has not yet been implemented (or the sentence is out of domain). During rule generation, the words/glosses ratio condition was implemented in order to direct the translation process to the statistical approach. The ratio threshold was fixed at 0.9.

4.3.2. Example-based strategy

An example-based translation system uses a set of sentences in the source language (sign language) and their corresponding translations in the target language (Spanish), and translates other similar source language sentences by analogy (Brown, 2002). In order to determine whether one example is similar enough for the text 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 text to be translated, in order to consider that it generates the same target sentence. In this case, the heuristic distance considered is the well known Levenshtein distance (LD) divided by the number of signs in the sentence to be translated (this distance is represented as a percentage). 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.

[11] http://www.nist.gov/speech/tests/mt/.

Please cite this article in press as: San-Segundo, R., et al. Spoken Spanish generation from sign language. Interact. Comput. (2009), doi:10.1016/j.intcom.2009.11.011

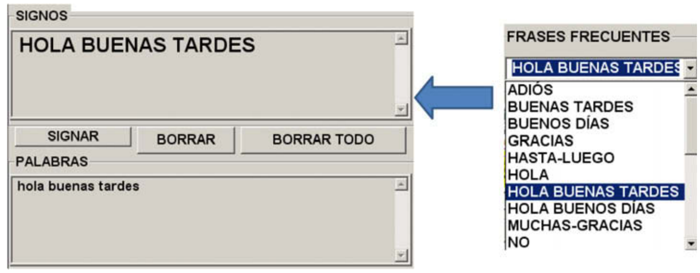

Fig. 16. Selection of a frequent sign sequence. In this case the selected sentence has been HOLA BUENAS TARDES (HELLO GOOD AFTERNOON).

Table 5
Results summary for rule-based and statistical approaches.

	WER	PER	BLEU	NIST
Statistical approach				
Phrase-based	39.94	33.24	0.4834	6.314
SFST-based	35.12	37.19	0.6402	6.930
Rule-based approach	24.52	19.43	0.6421	7.813

In order to make the examples more effective, it is possible to generalize them, so that more than one string can match any given part of the example. When indexing the example corpora, and before matching a new input against the database, the system tags the input by searching signs and phrases included in several class lists, and replacing each occurrence with the appropriate class name. There is a text 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, four classes were used: $NUMBER, $PROPER_NAME, $MONTH and $WEEK_DAY.

4.3.3. Rule-based strategy

In this strategy, the translation process is carried out in two steps. In the first one, every sign is mapped to one or several syntactic–pragmatic categories (categorization). After that, the translation module applies different rules that convert the tagged signs into words by means of grouping concepts or words (generally named blocks) and defining new words (San-Segundo et al., 2008). These rules are defined by an expert and can define short and large-scope relationships between concepts or signs. At the end of the process, the block sequence is expected to correspond to the word sequence resulting from the translation process.

The rules are specified in a proprietary programming language consisting of a set of primitives. The rule-based translation module implemented contains 129 translation rules and uses 10 different primitives. The evaluation measurements are presented in Table 5.

4.3.4. 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.[12] The phrase model has been trained using the GIZA++ software Och and Ney, 2000 (for calculating the alignments between signs and words) and the phrase extractor proposed by Koehn et al. (2003). The Moses decoder[13] was 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-g language model needed by Moses, the SRI language modelling toolkit has been used (Stolcke, 2002).

For the second method, the translation model consists of an SFST made up of aggregations: subsequences of source and target words aligned. The SFST is obtained from the word alignment (obtained with GIZA++) using the Grammatical Inference and Alignments for Transducer Inference (GIATI) algorithm (Casacuberta and Vidal, 2004). The SFST probabilities are also trained from aligned corpora. The software used in this paper has been downloaded from.[14]

In order to evaluate the different modules, the corpus (including only sentences pronounced by users) was divided randomly into three sets: training, development and test, carrying out a round-robin evaluating process. Table 5 summarizes the results for rule-based and statistical approaches: WER, PER, BLEU and NIST.

For this corpus the SFST-based method is better than the phrase-based method. For the summative evaluation presented in Section 4.6, statistical models were trained with the whole database. Both statistical alternatives were incorporated (phrase-based and SFST-based strategies), although only the SFST-based one was used for the summative evaluation because of its better performance.

4.4. Text to speech conversion

The text to speech (TTS) converter used was a commercial one: Loquendo (see footnote 10). Loquendo TTS is speech synthesis mark up language (SSML) compatible. SSML is a standard for generating content to be spoken by a speech synthesis system.[15] It is a W3C recommendation (see W3C press release and testimonials issued on September 8th 2004 or the full specification at[16]). SSML is an XML-based mark-up language, which is aimed at controlling text to speech conversion. The system presented in this paper used the "Jorge" Castilian voice though the Microsoft speech API.[17]

4.5. System scalability and its applicability to other domains

Creating scalable natural language interfaces is a very hard task that requires an important amount of resources (even when there is no speech or gesture recognition involved). These resources are necessary to model task knowledge properly. In the system proposed in this paper, the main modelling requirement is language translation. When using statistical approaches for automatic language translation, it is necessary a big parallel corpus including a significant amount of sentences in source and target languages (Och and Ney, 2003; Mariño et al., 2006). On the other hand, when considering rule-based approaches, an expert interpreter has to spend a lot of time defining the rules of the system. LSE has a tiny

12 http://www.statmt.org.
13 http://www.statmt.org/moses/.
14 http://prhlt.iti.es/content.php?page=software.php.
15 http://www.w3.org/TR/speech-synthesis/.
16 www.w3.org/TR/speech-synthesis/.
17 http://www.microsoft.com/speech/speech2007/default.mspx.

Please cite this article in press as: San-Segundo, R., et al. Spoken Spanish generation from sign language. Interact. Comput. (2009), doi:10.1016/j.intcom.2009.11.011

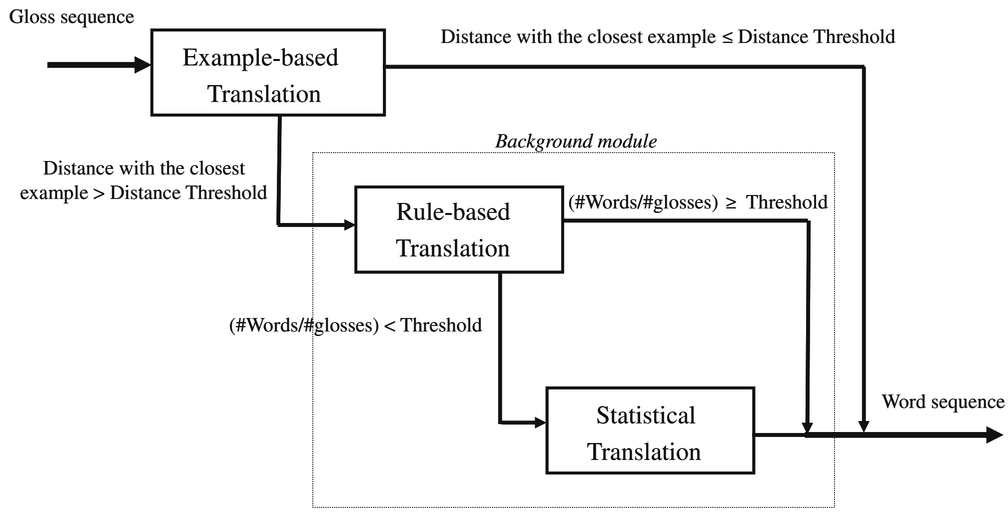

Fig. 17. Combination of three different translation strategies.

fraction of the resources that are available for English or even Spanish.[18]

Much of the effort for increasing the scope of a spoken dialogue system is focused on generating language resources. Once these resources exist, it is easier to extend the interface, specially, when most of the system functionality is data-driven, as it is our case (i.e. suggesting the next most likely signs, example-based and statistical translation modules). In the proposed system, there is no formal model of the driving licence renewal process that is hard-coded into the interface. So in that respect, considering other efforts for increasing the scalability of a spoken dialogue system (D'Haro et al., 2006; Bohus and Rudnicky, 2009), the interface is actually highly scalable for a natural language application.

This section describes the main aspects that must be considered when increasing the scope of the system (being able to translate more sentences) or when applying this system to another domain. The necessary changes affect two of the three modules that make up the system: visual interface and language translation. The text to speech conversion module works well for any sentence in Spanish, so it is not necessary to introduce any change in it.

For the visual interface, it is necessary to update the following aspects:

- The list of glosses considered in the search. For adding a new sign to the interface, the system needs to have a new file in a specific path, named with the sign gloss: i.e. CAR.txt. This file contains the sign description in SiGML, which must be represented by the avatar. This description can be generated using the sign editor presented in Section 3. When a new file is detected in the path, the interface updates the search with a new gloss (file name) and the avatar can represent it. From the SiGML, the system can obtain the corresponding HamNoSyS sign description. This task is the most time-consuming because every sign file must be generated by hand. For example, for generating 715 signs, one person working for 1 month was necessary. It is true that when one sign has been generated, it can be reused in any domain. On the other hand, the time and resources for generating a sign are lower compared to the option of recording a video for every sign. Section 3 has described the process for sign generation. In this process, the new Editor developed, with the possibility of translating SEA (available in the normative dictionary DILSE III developed by Fundación CNSE) into HamNoSys, has reduced the developing time by approximately 50%.
- The probabilities of the prediction tool should also be updated. In this case, a set of sign sequences (gloss sequences), similar to the set obtained in the parallel corpus in Section 3 is necessary. These probabilities are estimated automatically. For obtaining the corpus, three steps were necessary: the first was the collection of the Spanish sentence which took 1 month. Secondly, these sentences were revised by one person for correcting typing error (1 week). The last step was translating Spanish sentences into LSE (in glosses and videos). For this step, it was necessary for two people to work for 1 month.

In the case of the language translation module, it is necessary to update these components:

- For the example-based translation module, the examples of the database. These examples consist of Spanish sentences and their corresponding translation (a parallel corpus as described in Section 3).
- The rule-based translation module would need to develop new rules for translating new sentences. This is a time-consuming task because an expert must develop the rules by hand. Some of these rules (approximately 40%) are general translation rules and can be used in other domains, but there are a lot of them specific to this domain. In order to give an idea, the rules used in the system proposed were developed by one person over 3 weeks.
- For the statistical translation, it is necessary to update the translation models: these models are obtained automatically from a parallel corpus (Section 3).

It is possible to conclude that many aspects can be updated automatically from a parallel corpus, including sentences (in Spanish and LSE) related to the domain: except SiGML sign specifications and translation rules.

In order to increase the system adaptability, it will be necessary to work on the two most demanding tasks: sign and translation rule generation. In the case of sign generation, this paper (in Section 3) has described a new editor for sign specification including an SEA-HamNosys tool. With this editor it has been possible to reduce the time for sign generation by approximately 50%. For the case of translation rules, the authors are considering working on

[18] http://www.cervantesvirtual.com/seccion/signos/psegundonivel.jsp?conten=presentacion.

Table 6
Objective measurements for evaluating the Spanish generator from sign-writing.

Agent	Measurement	Value	User reading level	
			High	Medium–low
System	Translation rate (%)	98.0	98.1	97.9
	Average translation time (s)	0.001	0.001	0.001
	Average time for text to speech conversion (s)	1.7	1.7	1.7
	Cases using example-based translation (%)	91.9	92.2	91.6
	Cases using rule-based translation (%)	8.1	7.7	8.5
	Cases using statistical translation (%)	0.0	0.0	0.0
	Time for gloss sequence specification (s)	18.5	17.5	19.5
	# of clicks for gloss sequence specification (clicks)	8.6	7.6	9.5
	Time for gloss specification (s)	7.4	6.9	7.8
	# of clicks for gloss specification (clicks)	3.3	3.0	3.7
	# of glosses per turn	2.5	2.5	2.5
	Percentage of utility use			
	– List of glosses (%)	50.4	47.4	52.4
	– Search by Hamnosys (%)	0.0	0.0	0.0
	– List of proposed next glosses (%)	48.1	50.7	47.6
	– List of last used glosses (%)	1.5	2.6	0.0
	– Date (%)	0.0	0.0	0.0
	– Time (%)	0.0	0.0	0.0
	– Spelling (%)	0.0	0.0	0.0
	# of use of the most frequent sign sequences per dialogue	0.4	0.7	0.0
	# of user turn per dialogue	4.0	3.1	5.0
	# of dialogues	48	26	22
	– Scenario 1: having all the necessary documents	9	5	4
	– Scenario 2: no having identity document	8	4	4
	– Scenario 3: no having photo missing	8	4	4
	– Scenario 4: no having medical certificate	8	5	3
	– Scenario 5: needing to fill some information	8	4	4
	– Scenario 6: wanting to pay with cash	7	4	3

the statistical translation module in order to increase its performance. The statistical translator and the rule-based translator are both used as a background translation module in Fig. 17. If it is possible to achieve the same performance with both systems, the rule-based translation could be removed from this structure, avoiding the need of generating new rules when adapting the system to a new domain.

4.6. Summative evaluation and discussion

The spoken Spanish generator was evaluated for complementing a speech into LSE translator for driver's licence renewal. The speech-LSE system translates the government employee's explanations while the spoken Spanish generator developed in this work helps deaf users to ask questions. This section includes a detailed description of the evaluation carried out in the local traffic office in Toledo for the renewal of the driver's licence. In the evaluation, government employees, and Deaf from Madrid and Toledo were involved.

For renewing the driver's licence (DL) at the Toledo traffic office, users need to present a completed application form, a cash fee of 22 Euros, and a documentation including the identification document, the old DL, a medical certificate and a photo. The new DL is sent by mail in the next 3 months. For driving during this time, the user receives a provisional DL. For the evaluation, the whole user-officer interaction was carried out in the same desk.

The evaluation was carried out over 2 days. On the first day, a 1-h talk, about the project and the evaluation, was given to government employees (two people) and users (10 people) involved in the evaluation. Half of the users evaluated the system on the first day, leaving the other half for the next day. Six different scenarios were defined in order to specify real situations: in one scenario, the user simulated having all the necessary documents, three other scenarios in which the user simulated not having one of the documents: Identity document, photo or the medical certificate, one scenario where the user had to fill some information in the application form, and finally, a scenario in which the user wanted to pay with credit card (only cash is accepted according to existing paying policy). The sequence of scenarios was randomly selected for each user (different order between users).

Ten deaf users interacted with two government employees at the Toledo traffic office using the developed system. These ten users (six males and four females) tested the system in almost all the scenarios described previously: 48 dialogues were recorded (12 dialogues were missing because several users had to leave the evaluation session before finishing all the scenarios). The user ages ranged between 22 and 55 years old with an average age of 40.9 years. Half of the users (5 of 10 users) said that they worked with a computer every day and the other half use a computer a few times per week. Only half of them had a high reading level and they are used to read glosses for sign sequence specification. There is a strong correlation between these user characteristics: the users who work with a computer every day generally have a medium–high reading level and a better habit of using glosses.

The evaluation includes objective measurements from the system and subjective information from both user and government employee questionnaires. A summary of the objective measurements obtained from the system are shown in Table 6. These measurements have been obtained using a capturing software (Camtasia studio 6[19]) and a detailed log generated by the system.

As is shown, the good translation rate and the short translation time make it possible to use this system in real conditions. Related to the translation process, the example-based strategy has been selected in most of the cases. This behaviour shows the reliability of the corpus collection: the most frequent user questions were taken down, obtaining a very good representative corpus for this kind of dialogue. In this evaluation, it was not necessary to use the statistical translation module because the rule-based module could deal with the translation of gloss sequences not contained in the corpus. If the system is used with a scope bigger than the corpus collection,

[19] http://camtasia-studio.softonic.com/.

Please cite this article in press as: San-Segundo, R., et al. Spoken Spanish generation from sign language. Interact. Comput. (2009), doi:10.1016/j.intcom.2009.11.011

the example-based translation would be used less and the rule-based or the statistical strategies would have to deal with new sentences, thus decreasing the system performance. For future work, the authors are considering evaluating the system in a bigger scenario to see how rule-based and statistical translation can deal with new sentences. If the new scope is an extension of the current domain, the decrease in performance is expected to be very low and it can be taken on. On the other hand, if the new scope is a new domain the performance would be very low: it is necessary to establish a new corpus collection (and a new rule generation process in the case of using the rule-based translation).

The user needed less than 20 s to specify a gloss sequence using the interface. This time is short considering that the user had only few minutes to practice with the visual interface before the evaluation (the little use of the most frequent sign sequences list reveals a limited experience with the interface) and the final interface is a bit complex. With more time for practising, this time would be reduced. Of course, this time is higher compared to the time needed by an automatic sign recognition system (3–5 s) (Sylvie and Surendra, 2005) but the performance is considerably better because sign recognition technology is not yet mature enough to be accepted by the Deaf.

Regarding the different utilities for gloss specification, it is important to comment that the alphabetical list of glosses was used in around 50% of the times (this result is understandable because it is necessary, at least for the first gloss, to select one letter and pick up one gloss from the alphabetical list of glosses). The use of one gloss predicted by the system from a popup menu (based on the most probable sign sequences) has been used in more than 48% of the times. This utility has been very useful in order to speed up the gloss sequence specification process. The date, time and spelling utilities have not been used because it was not necessary to specify a date, time or proper name in the proposed scenarios.

Table 6 also shows objective measurements for two different user groups depending on their reading level: high and medium–low levels. Although, the differences are not very big, it is possible to see how users with a lower reading level, need more time and clicks for specifying the sign sequence.

The subjective measures were collected from questionnaires filled in by both government employees and deaf users. They evaluated different aspects of the system giving a score of between 0 and 5 (strongly disagree–strongly agree). The questionnaires were designed by the original group of experts. Designing questionnaires for the Deaf is complex because of three issues:

1. The first problem is deciding on the language for asking the different questions: LSE (using videos) or written Spanish. In this case, the decision was to present the questions in Spanish with translation in LSE (glosses) and having two interpreters for solving any questions.
2. The second problem was to decide the aspects to evaluate and question design. The first idea was to reuse questionnaires developed for evaluating Speech-based applications (Möller et al., 2007) or Human-Computer Interaction (HCI) systems (Brooke, 1996). Immediately, experts in LSE (Deaf) reported the problem that for many of the concepts and words used in these questionnaires have no translation into LSE, so many of these concepts would be difficult for the Deaf to understand. (i.e. questionnaires items that are difficult to translate: *I thought there was too much inconsistency in this system* or *I found the various functions in this system were well integrated*). Because of this aspect, the expert panel decided to reduce the number of questions, designing them based on tangible aspects (easier to explain with examples).
3. Finally, the third aspect was the scale used: number of levels and the names for the different levels. For the number of levels, the expert panel decided to define an even number (six in this case) eliminating the neutral level and forcing the user to decide. One reason is because this neutral level is the most common refuge when a user does not understand one of the questions very well. Forcing a user to decide causes this user to ask the interpreter more questions in order to understand all the details. A second reason was that it is very difficult to find deaf users for evaluating this kind of system and the authors wanted to obtain the best feedback with a small number of users. As regards the label for the different levels, the final decision was to specify six numerical levels providing information for the level 0 and 5 (strongly disagree, strongly agree). Defining labels for all the levels is a difficult problem because the differences between consecutive levels cannot always be described properly using LSE. There is a high probability that the nuances were not perceived by a deaf person, while a numerical scale is easier to understand.

The main results are presented in Table 7.

The government employees have assessed both speech intelligibility and naturalness well. The system uses the male voice "Jorge" of Loquendo TTS. The users gave a reasonable score to the visual interface but they also reported some problems:

- The first comment is that there was not enough time to practice with the system. It is not possible to learn all utilities in just a few minutes. This aspect can be also associated to the interface complexity.
- Another significant problem is that gloss notation is not yet standardized enough for LSE: one sign can be represented by two glosses (i.e. the sign "YO" ("I") can be represented by "YO" or by "MI" (me)).

In order to solve the second problem, some users suggested incorporating a sign selection mechanism based on images or gifs (small animations). This possibility was considered during the interface design but it was ruled out because the reasons presented in Section 4.1: Deaf associate this kind of interface to children, mentally disabled people or people with learning disabilities. It is very important to avoid this kind of association because, traditionally, deafness has been associated to people with learning problems but this is not true. The use of sign languages defines Deaf as a linguistic minority, with learning skills, cultural and group rights similar to other minority language communities (Key and Allsop, 1997; Leeson et al., 2005).

Another alternative for the gloss notation ambiguity is using SEA or HamNoSys for writing the sign sequence. This alternative was discarded during the design because these two notation sys-

Table 7
Subjective measures for evaluating the Spanish generator from sign-writing.

Agent	Evaluation measurement	Mean (0–5)	Standard deviations (0–5)
Government employee	The speech is intelligible	4.0	0.0
	The speech is natural	3.0	0.0
	I had use the system in absence of a human interpreter	4.0	0.7
	Global assessment	**4.0**	0.7
Deaf users	The system is fast	2.6	1.2
	The system is easy to learn	2.2	2.0
	The system has enough glosses	2.7	1.3
	I had use the system in absence of a human interpreter	2.1	1.7
	Global assessment	**2.5**	1.3

tems are very useful for sign description but they are not well known by the Deaf.

In order to report more information on the user evaluation, Fig. 19 shows the distribution of the number of users versus the overall assessment provided. This analysis reveals two different perceptions about the use of new technologies. Analysing the impact of the user reading level on the subjective evaluation, it is shown that users with a lower reading level evaluated the system with a worse score for all the items: three of the four users that assess the interface with a global score of one (0–5 scale) have a medium–low reading level. Table 8 shows subjective measurements for two different user groups depending on their reading level: high and medium–low levels.

Table 9 shows the Spearman's correlation between deaf user evaluation and their background: computer experience, confidence with written Spanish and using glosses. This table also includes *p*-values for reporting the correlation significance. Because of the very low number of data and the unknown data distribution, the 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.

As it is shown, all the evaluation measurements correlate positively with all user characteristics, except with "the system has enough glosses". In this case, the correlation is close to zero (uncorrelated aspects). Generally, a better computer experience, a better confidence with written Spanish, and a better confidence using glosses for sign-writing, are correlated to a better system evaluation. The confidence with written Spanish and the confidence with glosses are the characteristics that better correlate with the user evaluation for all measurements. As regards the evaluation measurements, the use of the system in the absence of a human interpreter and the overall evaluation, have been the measurements with the higher correlation with the user background.

Finally, an interesting suggestion from users was to implement a mechanism to synchronise both systems: the Spanish into sign language translation system and the spoken Spanish generator from LSE. The main idea is to avoid the government employee speaking while the user is compiling the gloss sequence. The position of both government employee (with the microphone) and user is shown in Fig. 18.

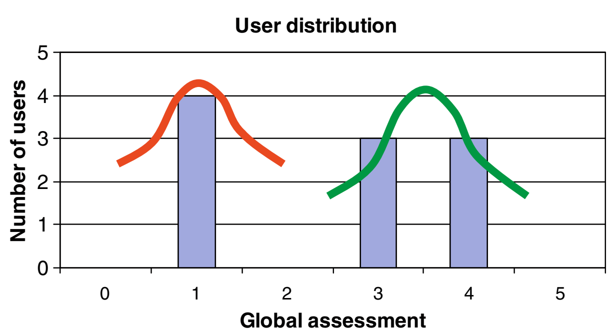

Fig. 19. Distribution of users versus global assessment.

Table 8
Subjective measures depending on the user reading level.

Agent	Evaluation measurement	User reading level	
		High	Medium–low
Deaf users	The system is fast	3.2	1.8
	The system is easy to learn	3.2	0.8
	The system has enough glosses	2.7	2.7
	I had use the system in absence of a human interpreter	3.0	0.8
	Global assessment	**3.2**	**1.5**

5. Main conclusions

This paper has presented the first spoken Spanish generator from sign-writing of Spanish sign language (LSE: Lengua de Signos Española). This system consists of an advanced visual interface where a deaf person can specify a sequence of signs in sign language, a language translator (for generating the sequence of words in Spanish), and finally, a text to speech converter. The visual interface allows a sign sequence to be developed using several sign language alternatives. The analysis and the design process of this interface have been described in detail, reporting very useful information for designing visual interfaces for the Deaf. For language translation, three strategies were developed and combined to implement the final version of the language translator module. The paper also includes an analysis about the system scalability: what elements should be updated when changing the domain and an estimation of the effort required.

In the summative evaluation, the system performed very well in language translation (2.0% word error rate). The users gave a reasonable positive score but some problems related with the time for practising and with the level of gloss standardization were reported. The user needed less than 20 s to specify a gloss sequence using the interface. This time is low considering that the user had only few minutes to practice. This time is higher compared to the time needed by an automatic sign recognition system (3–5 s) but the performance is considerably better.

Secondly, this paper also has described the first Spanish-LSE parallel corpus for language processing research focusing on specific domains: The renewal of the identity document and driver's license. This corpus includes 4080 Spanish sentences translated into LSE. This corpus also contains a sign database including all sign descriptions in several sign-writing specifications: Glosses, HamNoSys and SEA: Sistema de Escritura Alfabética. This sign database includes all signs in the parallel corpus and signs for all the letters (necessary for word spelling), numbers from 0 to 100, numbers for time specification, months and week days. The sign database has been generated using a new version of the eSign editor. This new version includes a grapheme to phoneme system for Spanish and a SEA-HamNoSys converter. Most of the signs generated (around 70%) were included in the Fundación CNSE dictionary so the SEA description was available and the HamNoSys notation

Table 9
Analysis of correlations between deaf user evaluation and their background.

EVALUATION MEASUREMENT	Computer experience	Confidence with written Spanish	Confidence with glosses
The system is fast	0.18 (p>0.500)	**0.62 (p=0.080)**	**0.62 (p=0.080)**
The system is easy to learn	**0.48 (p=0.135)**	**0.62 (p=0.080)**	**0.62 (p=0.080)**
The system has enough glosses	-0.12 (p>0.800)	0.00 (p>0.800)	0.00 (p>0.800)
I'd use the system in absence of a human interpreter	**0.47 (p=0.140)**	**0.69 (p=0.051)**	**0.69 (p>0.051)**
GLOBAL assessment	**0.37 (p=0.230)**	**0.64 (p=0.060)**	**0.64 (p>0.060)**

Please cite this article in press as: San-Segundo, R., et al. Spoken Spanish generation from sign language. Interact. Comput. (2009), doi:10.1016/j.intcom.2009.11.011

Fig. 18. Different photos of the evaluation process at the Toledo Traffic Office.

(needed to represent the sign with the eSIGN avatar) was generated automatically using the SEA-HamNoSys converter. In all these cases, the SEA-HamNoSys conversion tool has been very useful because it has reduced significantly the design time, by approximately 50%.

6. Future work

In order to define the future work, the same group of experts that designed the interface have analysed the evaluation results. There is a lot of work to do in order to improved the interface. Special attention must be paid on deaf users with low experience using the computer, or low confidence with written Spanish or glosses. From this analysis, this group has proposed new changes for the next version of the interface:

- The first change is to remove the searching tool based on HamNoSys from the interface. From the results, it has been demonstrated that this standard is not known by the Spanish Deaf. Perhaps, this tool can be incorporated into a submenu, but in any case, it can appear in the main window.
- The proposal for the list of the last five introduced glosses is also to remove it from the main window, leaving more space for other tools.
- It is true that using pictures for sign search has several problems as reported in 4.1, but it is possible to incorporate icons into some of the tool buttons, specifically in five buttons: SIGNAR (for representing the whole sign sequence), BORRAR (to delete the last sign), BORRAR TODO (to delete the whole sign sequence), DELETREO (to change to the spelling state), and HABLA (for speaking the Spanish sentence).
- The sponsor logos should be included only in promotional events but for real use, they must be removed (or drastically reduced in size).

Another aspect considered to improve the interface is to perform experiments with deaf users to see how they would classify the key concepts involved in the application. These experiments can be done using the Card Sorting methodology[20] (Akerelrea and Zimmerman, 2002).

Acknowledgements

The authors want to thank the essential sign language information on government networks (eSIGN) consortium for permitting the use of the eSIGN editor and the 3D avatar in this research work. The authors want to thank discussions and suggestions from the colleagues at GTH-UPM and Fundación CNSE. This work has been supported by Plan Avanza Exp. No.: PAV-070000-2007-567, ROBONAUTA (MEC ref: DPI2007-66846-c02-02) and SD-TEAM (MEC ref: TIN2008-06856-C05-03) projects. Authors also want to thank Mark Hallett for the English revision and the reviewers for their valuable comments.

[20] http://www.stcsig.org/usability/topics/cardsorting.html.

References

Akerelrea, C., Zimmerman, D., 2002. A group card sorting methodology for developing informational web sites. In: Proceedings of the 2002 IEEE Professional Communications Conference, pp. 437–445.

Bohus, D., Rudnicky, A., 2009. The RavenClaw dialog management framework: architecture and systems. Comput. Speech Lang. 23 (3), 332–361.

Brooke, J., 1996. SUS: a quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (Eds.), Usability Evaluation in Industry. Taylor & Francis, London.

Brown, R.D., 2002. Example-Based Machine Translation, Tutorial at Conference of the Association for Machine Translation. AMTA-2002.

Bungeroth, J., Stein, D., Dreuw, P., Zahedi, M., Ney, H., 2006. A German sign language corpus of the domain weather report. In: 5th International Conference on Language Resources and Evaluation, Genoa, Italy, May 2006.

Bungeroth, J., Stein, D., Dreuw, P., Ney, H., Morrissey, S., Way, A., van Zijl, L., 2008. The ATIS sign language corpus. In: International Conference on Language Resources and Evaluation (LREC), Marrakech, Morocco, May 2008.

Campr, P., Hruz, M., Trojanová, J., 2008. Collection and Preprocessing of Czech Sign Language Corpus for Sign Language Recognition. Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia in Pilsen. LREC 2008.

Casacuberta, F., Vidal, E., 2004. Machine translation with inferred stochastic finite-state transducers. Comput. Linguist. 30 (2), 205–225.

D'Haro, L.F., Córdoba, R., Ferreiros, J., Hamerich, S.W., Schless, V., Gladis, B., Schubert, V., Koesis, O., Igei, S., Pardo, J.M., 2006. An advanced platform to speed up the design of multilingual dialog applications for multiple modalities. Speech Commun. 48 (8), 863–887.

Dreuw, P., 2008. Visual modeling and tracking adaptation for automatic sign language recognition. In: International Computer Vision Summer School (ICVSS), Sicily, Italy, July 2008.

Dreuw, P., Neidle, C., Athitsos, V., Sclaroff, S., Ney, H. 2008a. Benchmark databases for video-based automatic sign language recognition. In: International Conference on Language Resources and Evaluation (LREC), Marrakech, Morocco, May 2008.

Dreuw, P., Forster, J., Deselaers, T., Ney, H., 2008b. Efficient approximations to model-based joint tracking and recognition of continuous sign language. In: IEEE International Conference Automatic Face and Gesture Recognition (FG), Amsterdam, The Netherlands, September 2008.

Dreuw, P., Stein, D., Deselaers, T., Rybach, D., Zahedi, M., Bungeroth, J., Ney, H., 2008c. Spoken language processing techniques for sign language recognition and translation. J. Technol. Disabil. 20, 121–133. ISSN 1055-4181.

Dreuw, P., Stein, D., Ney, H., 2009. Enhancing a sign language translation system with vision-based features. In: Special Issue Gesture Workshop 2007, LNAI, No. 5085, Lisbon, Portugal, January 2009, pp. 108–113.

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

Hanke, T., Popescu, H., 2003. Intelligent Sign Editor. eSIGN Project Deliverable. D2.3.

Herrero, A., 2004. Escritura alfabética de la Lengua de Signos Española, Universidad de Alicante, Servicio de Publicaciones.

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

Key, Allsop, 1997. Sign on Europe: A Study of Deaf People and Sign Language in the European Union. Centre for Deaf Studies, University of Bristol, Bristol.

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

Leeson, L., 2005. Signed Languages in Education in Europe – A Preliminary Exploration. Language Policy Division, Council of Europe, Strasbourg. <www.coe.int/lang>.

Leeson, L., Saeed, J., Macduff, A., Byrne-Dunne, D., Leonard, C., 2006. Moving heads and moving hands: developing a digital corpus of Irish sign language. In: Proceedings of Information Technology and Telecommunications Conference 2006, Carlow, Ireland.

Mariño, J.B., Banchs, R., Crego, J.M., Gispert, A., Lambert, P., Fonollosa, J.A., Costa-Jussà, M., 2006. N-gram-based machine translation. Comput. Linguist. Assoc. Comput. Linguist. 32 (4), 527–549.

Möller, S., Smeele, P., Boland, H., Krebber, J., 2007. Evaluating spoken dialogue systems according to de-facto standards: a case study. Comput. Speech Lang. 21 (1), 26–53.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., Lee, R.G., 2000. The Syntax of American Sign Language. MIT Press, Cambridge, MA, USA.

Och, J., Ney, H., 2000. Improved statistical alignment models. In: Proceeding of the 38th Annual Meeting of the Association for Computational Linguistics, Hongkong, China, October 2000, pp. 440–447.

Och, J., Ney, H., 2003. A systematic comparison of various alignment models. Comput. Linguist. 29 (1), 19–51.

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), Philadelphia, PA, pp. 311–318.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Henning, J., et al., 1989. Hamburg Notation System for Sign Languages – An Introductory Guide. International Studies on Sign Language and the Communication of the Deaf, vol. 5. Institute of German Sign Language and Communication of the Deaf, University of Hamburg.

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 Commun. 50, 1009–1020.

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

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

Su, H.-Y., Chiu, Y.-H., Wu, C.-H., Cheng, C.-J., 2007. Joint optimization of word alignment and epenthesis generation for Chinese to Taiwanese sign synthesis. IEEE Trans. Pattern Anal. Mach. Intell. 29 (1), 28–39.

Sylvie, O., Surendra, R., 2005. Automatic sign language analysis: a survey and the future beyond lexical meaning. IEEE Trans. Pattern Anal. Mach. Intell. 27 (6).

Tryggvason, J., 2004. VANESSA: A System for Council Information Centre Assistants to Communicate Using Sign Language. School of Computing Science, University of East Anglia.

Vogler, C., Metaxas, D., 2001. A Framework for recognizing the simultaneous aspects of ASL. CVIU 81 (3), 358–384.

von Agris, U., Schneider, D., Zieren, J., Kraiss, K.-F., 2006. Rapid signer adaptation for isolated sign language recognition. In: CVPR Workshop V4HCI, New York, USA, June 2006, pp. 159.

Wang, S.B., Quattoni, A., Morency, L.-P., Demirdjian, D., Darrell, T., 2006. Hidden conditional random fields for gesture recognition. In: CVPR, vol. 2, June 2006, pp. 1521–1527.

Wells, J.C., 1997. SAMPA computer readable phonetic alphabet. In: Gibbon, D., Moore, R., Winski, R. (Eds.), Handbook of Standards and Resources for Spoken Language Systems. Mouton de Gruyter, Berlin and New York (Part IV, Section B).

Yao, G., Yao, H., Liu, X., Jiang, F., 2006. Real time large vocabulary continuous sign language recognition based on OP/Viterbi algorithm. In: 18th ICPR, vol. 3, August 2006, pp. 312–315.

Zwitserslood, I., Verlinden, M., Ros, J., van der Schoot, S., 2004. Synthetic signing for the deaf: eSIGN. In: Proceedings of the Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment, CVHI 2004, 29 June–2 July 2004, Granada, Spain.

Please cite this article in press as: San-Segundo, R., et al. Spoken Spanish generation from sign language. Interact. Comput. (2009), doi:10.1016/j.intcom.2009.11.011