Methodology for developing a Speech into Sign Language Translation System in a New Semantic Domain

Verónica López-Ludeña[1], Rubén San-Segundo[1], Carlos González[2], Juan Carlos López[2] y José M. Pardo[1]

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

veronicalopez@die.upm.es

Abstract. This paper proposes a methodology for developing a speech into sign language translation system considering a user-centered strategy. This methodology consists of four main steps: analysis of technical and user requirements, data collection, technology adaptation to the new domain, and finally, evaluation of the system. The two most demanding tasks are the sign generation and the translation rules generation. Many other aspects can be updated automatically from a parallel corpus that includes sentences (in Spanish and LSE: Lengua de Signos Española) related to the application domain. In this paper, we explain how to apply this methodology in order to develop two translation systems in two specific domains: bus transport information and hotel reception.

Keywords: Methodology, New domain, Deaf people, Spanish Sign Language (LSE), Spoken Language Translation, Sign Animation

1 Introduction

Defining a methodology is an important aspect in order to develop human-computer interaction systems. A methodology ensures that quality requirements are reached (performance, time and cost). The usability of the developed system will depend on that quality and, as a result of a higher usability, there will be a better acceptance from deaf people. In this paper, we explain the methodology used in the CONSIGNOS project (Plan Avanza Exp Nº: TSI-020100-2010-489). This project aims to adapt a Spanish into Spanish Sign Language (LSE: Lengua de Signos Española) translation system, already developed in other specific domains (like driver's license renewing service), to different domains: bus transport information and hotel reception.

2 State of the art

Two of the most relevant projects in speech into sign language translation, ViSiCAST and eSIGN (Elliot et al., 2008), follow a specific methodology. The ViSiCAST project (Elliot et al., 2008), whose main goal was to produce adaptable communication tools allowing sign language communication, was structured to have three application-oriented work packages, each focusing on the technical issues in delivery for that specific application area, and two enabling technology work packages, focusing on virtual signing, sign language representation, and sign language synthesis from conventional textual sources. A further evaluation work package is concerned with eliciting feedback from deaf people at various stages within the development of the system.

The eSIGN project aimed to provide sign language on websites. The different tasks of this project are: development of tools needed for creating signed content; improvement of the signed output the avatar; creating the first information sites on the Internet with animated signing; content creation in all three partner countries; the further development of tools needed for creating signed content; further improvement of the signed output of the avatar and the user involvement and continued evaluation of their tools and the avatar's comprehensibility.

A similar Spanish project "Speech into Spanish Sign Language Translation for a personal service" (San Segundo et al., 2008; San Segundo et al., 2011) had the following developing tasks: linguistic study, system design, speech recognizer design, rule-based translation development, statistic translation development, sign generation, evaluation and documentation of the project.

Another example of sign language translation systems are: VANESSA (Voice Activated Network Enabled Speech to Sign Assistant) project (Tryggvason, 2004) that was part of eSIGN and that facilitates the communication between assistants and their deaf clients in UK Council Information Centres (CIC's) or similar environments. Other two main research projects that focus on sign language recognition are DICTA-SIGN (Hanke et al., 2010; Efthimiou et al., 2010) and SIGN-SPEAK (Dreuw et al., 2010a and 2010b). DICTA-SIGN aims to develop the technologies necessary to make Web 2.0 interactions in sign language possible. In SIGN-SPEAK, the overall goal is to develop a new vision-based technology for recognizing and translating continuous sign language into text.

There are many other examples of design, development and evaluation of Sign Language translation systems. In (Cox et al., 2003), it is described the design, implementation and testing of an experimental interactive translation system that aims to aid a deaf person in transactions in a Post Office and they also explain the evaluation of the system. The paper (Huenerfauth et al., 2008) describes an implementation and user-based evaluation (by native ASL signers) of a prototype ASL natural language generation system that produces sentences containing classifier predicates, which are frequent and complex spatial phenomena that previous ASL generators have not produced.

3 Methodology overview

This paper presents an adaptation of the Participatory Design methodology: one of the most used User-Centered Design approaches that follows the ISO standard Human-centered design for interactive systems: ISO 9241-210, 2010. Participatory design (known before as 'Cooperative Design') is a designing approach where all stakeholders (e.g. employees, partners, customers, citizens, and end-users) are involved actively in the design process. The main target is to guarantee that the final designed product meets their needs and is usable. This methodology consists of the following phases or tasks (**Fig. 1**):

- The requirement analysis (phase 1) will be undertaken with two Participatory Design workshops where end-users (Deaf people), researchers and developers will work together to define the system and user requirements.
- The data collection (phase 2) will be performed using preliminary prototypes.
- During technology adaptation (phase 3), users must supervise the process in order to guarantee that all user requirements are considered.
- The evaluation (phase 4): evaluation design, field evaluation and results analysis.

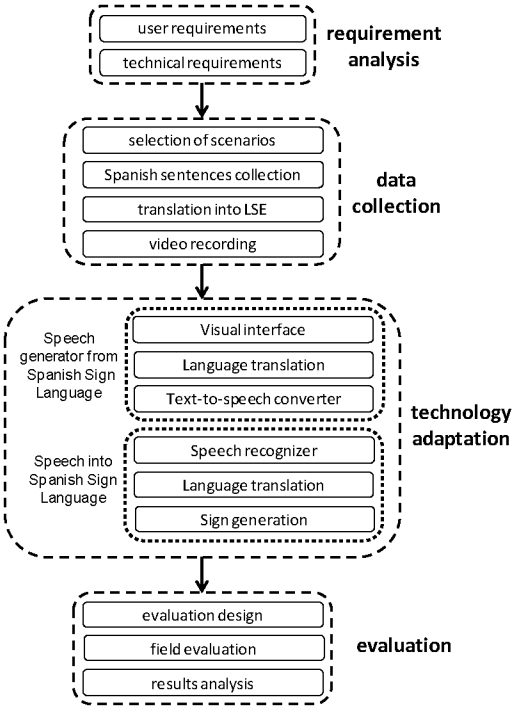

Fig. 1. Phases of system adaptation to a new application domain

The following sections explain the application of each phase.

4 Requirement analysis (Phase 1)

This section describes the process to analyze and to collect the main requirements of deaf users in relation to the development of an advanced communication system that will facilitate the communication between deaf and hearing people.

4.1 User requirements

According to the Survey of Disability, Personal Autonomy and Dependency Situations (EDAD, 2008) from INE (Spanish Statistic Institute), there are 1,064,100 deaf people in Spain. Deafness brings about significant communication problems: most deaf people have problems when expressing themselves in oral languages or understanding written texts. Their communication barriers are the main cause of the fact that 47% of deaf population has not studies or even is illiterate (INE -Spanish Statistic Institute- 1999 y MEC -Science and Education Ministry- 2000/2001).

All these aspects support the need to generate new technologies in order to develop automatic translation systems for converting this information into sign language.

First of all, it is necessary to define the scenarios of the system that we need to develop. In the case of our translation system, two different scenarios have been considered that correspond to personal attention services. In these two scenarios, two independent demos will be developed: the first scenario is the personal attention service that the EMT (Municipal Transport Enterprise) provides to users. The second is the personal attention service in a hotel reception. For each one of these services, the next steps have been followed for the analysis and definition of requirements:

- Visit to the facilities of the personal attention service site and interviews with employees. 3-4 interviews were carried out with the employees of EMT and the hotel in order to define the application domain for the demos.
- General description of service; the most requested information by users; service schedule and distribution of consultations per schedules; description of information offered to users; kind of users, etc.

4.2 Technical requirements

An important challenge of the project is to achieve a minimum level of quality of the technology, because the usability of the developed systems will depend on that quality. The following technologic requirements are defined:

- Speech recognition system will demand a recognition rate higher than 90% in the selected application domain. If that rate is not reached with speaker-independent models, an adaptation process will be performed to each speaker involved in the evaluation in order to guarantee that rate.
- It is also necessary a translation rate higher than 90% for the application domains targeted in the project. In order to obtain that translation rate, several translation strategies will be analyzed and combined.

– Finally, it is necessary an intelligibility of the avatar of 90% when representing the signs: recognition rate of deaf people. In order to obtain this intelligibility, the sign generation uses techniques based on inverse kinematics and semi-automatic movement capture that lets to obtain more realistic movements. This approximation requires more time for the dictionary generation, but it is more realistic.

In order to guarantee these technological requirements, it will be required a Spanish-LSE parallel corpus with a significant number of sentences in the application context. An approximated estimation for each service would be around 500 sentences containing less than 1,000 Spanish words and less than 200 signs in Spanish Sign Language.

5 Data collection (Phase 2)

5.1 Sentence recording process

In order to obtain the database, the next four steps are necessary:

– Selection of scenarios and application domain. The application domains were selected based on a guided visit to the installations and a set of interviews with the employees that offer the services
– Collection of sentences in Spanish. Once we defined the scenarios or application domains, researchers went to different point of service and collected sentences that were pronounced by both users and employees. This collection took several days.
– Translation of Spanish sentences into LSE (Spanish Sign Language). All Spanish sentences were translated by Deaf people that know Spanish language.
– Video recording. All sentences were represented by LSE experts.

5.2 Database statistics

	EMT		**HOTEL**	
	Spanish	**LSE**	**Spanish**	**LSE**
Total number of pair of sentences	600	600	499	499
Different sentences	523	362	493	333
Total number of words or glosses	3801	2800	3850	2780
Vocabulary	627	303	734	350

Table 1. Main statistics of the corpus

Table 1 shows the main characteristics of the corpus. The "vocabulary" row indicates the number of different words or glosses. There are 523 different sentences from the EMT information service and 493 different sentences from different situations in a

hotel reception, with corresponding translation into LSE. There are fewer different LSE sentences than Spanish sentences, because different Spanish sentences can have the same LSE translation.

6 Adapting the speech and language technologies to a new domain (Phase 3)

This section describes the main aspects that must be considered when increasing the system scope or when applying this system to other domain.

6.1 Speech into Spanish Sign Language translation system

The Spanish into LSE translation system converts natural speech sentences into LSE sentences signed by an avatar. This system is made up of three modules (**Fig. 2**). The first one is a speech recognizer that converts natural speech into a sequence of words (text). The second one is a natural language translation module that converts a word sequence into a sign sequence. And the third one is an avatar that represents the signs.

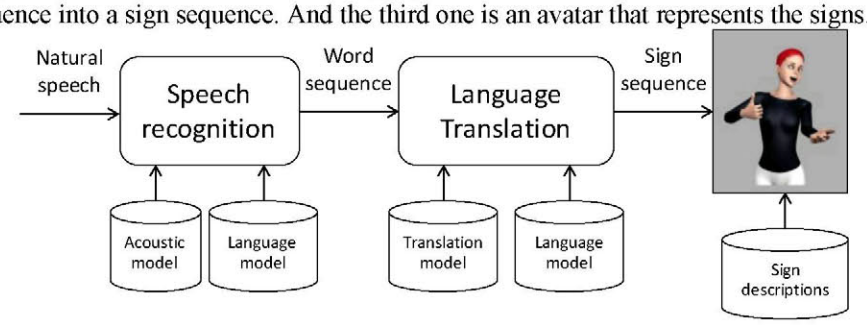

Fig. 2. Speech into Spanish Sign Language translation system

In order to adapt this system, it is necessary update the three modules.

- **Speech recognizer**

Vocabulary and language models of the speech recognition system must be updated using the Spanish sentences collected for each application domain. The speech recognizer includes an acoustic adaptation module for adapting the acoustic models to a new specific environment, a new speaker, or a new Spanish accent. Also, when generating automatically the vocabulary and a language model for the speech recognizer, a new module has been included for introducing source language variants, increasing the speech recognizer flexibility.

- **Language translation module**

The translation module has a hierarchical structure divided into two main steps. In the first step, an example-based strategy is used to translate the word sequence. And there

is also a background module that translates the sentence if the previous translation is not good enough. For the background module, a combination of rule-based (where a set of translation rules, defined by an expert, guides the translation process) and statistical translators (phrase-based translator and Finite State Transducers) has been used.

For the example-based translation module, it is necessary to update the examples of the database. These examples consist of Spanish sentences and their corresponding translation (the parallel corpus described in section 5).

The rule-based translation module would need to develop new rules for translating new sentences. It is a time-consuming task because an expert must develop the rules by hand. Some of the rules (approx. 40%) are general translation rules and they can be used in other domains, but a lot of them are specific for this domain. The proposed rules used in the system were developed by one person during 3 weeks.

For the statistical translation, it is necessary to update the translation models: these models are obtained automatically from a parallel corpus (section 5). Because the rule-based strategy is the most demanding task, the statistical translation strategy incorporates a new pre-processing module (López-Ludeña et al, 2011) that permits to increase its performance, and to replace the rule-based translation strategy.

- **Sign representation**

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

In order to increase its adaptability, the sign animation module includes a new version of the sign editor that incorporates new options (like predefined positions and orientations) for reducing significantly the sign specification time.

6.2 Spanish generator from Spanish Sign Language

For the Spanish generator from LSE (**Fig. 3**), the necessary changes affect two of the three modules composing 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 on it.

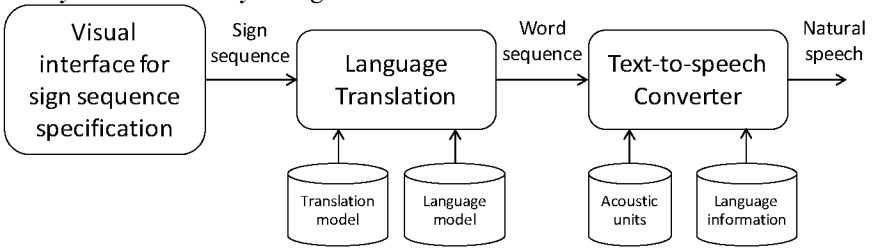

Fig. 3. Spanish generator from Spanish Sign Language

- **Visual interface.**

It is necessary to update the list of glosses considered in the search. When a deaf user specifies a certain gloss, an avatar represents it in order to allow deaf user checking that the selected gloss is appropriate. In order to add a new sign to the interface, the system needs to have a new file in a specific path, named with the sign gloss: e.g. CAR.txt. This file contains the sign description, mandatory to be represented by the avatar. This description can be generated using a sign Editor. 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.

Generating the signs is the most time-consuming task, because every sign file must be generated by hand. For example, for generating 715 signs (needed in these two domains), it was necessary one person working during 2 month. 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. A new Editor has been developed reducing the developing time more than 50% approx. by using techniques based on inverse kinematics and semi-automatic movement capture that lets to obtain more realistic movements.

- **Translation module**

It is the same module as commented in the previous section, but in opposite translation direction. So, the update is the same.

7 Evaluation (Phase 4)

7.1 Design and field evaluation

In order to carry out a system evaluation it is necessary to take into account several aspects. The evaluation should be carried out in at least two days and around twelve people should be involved: 2 employees of each application domain and 10 deaf users. On the other hand, it is necessary to define different scenarios (around five or six) in order to test real situations. For instance, asking for different kinds of information or service or simulating different kind of problems. The ten users should interact with the two employees and all of them should test the system in almost all the scenarios. This way, finally we should have around 50 dialogues.

It is important to collect the age's range of deaf users, their understanding level of written Spanish, their habit of using glosses for sign sequence specification and their practice with computers.

7.2 Objective measurements

Evaluation must include objective measurements from the system and subjective information from both user and employee questionnaires.

The objective measurements can be obtained using a capturing software (Camtasia Studio 6: http://www.techsmith.com/camtasia.html) and a detailed log generated by the system. These measurements should include the next aspects: speech recognition rate, translation rate, average translation time, average time for text to speech conversion, percentage of cases using example-based translation, percentage of cases using each translation technique (example-based, rules-based or statistical), time for gloss sequence specification, number of user turn per dialogue, number of dialogues, etc.

7.3 Subjective measurements

Subjective measurements can be collected from questionnaires filled by both employees and deaf users. In these questionnaires, they could evaluate different aspects of the system giving a score between 0 and 5. Designing questionnaires for Deaf is an interesting aspect with several problems that it is important to deal with. A group of experts have to design the questionnaires taking into account the following aspects:

- To decide the language for asking the different questions: LSE (using videos) or written Spanish. A good solution would be to present the questions in Spanish with translation in LSE (glosses) and having two interpreters for solving any question.
- To decide the aspects to evaluate and design questions. Experts in LSE (Deaf) report problems with concepts/words used in questionnaires developed for evaluating Speech-based applications (Möllera et al, 2007) or Human-Computer Interaction (HCI) systems (Brooke, 1996), because these questionnaires have not translation into LSE, so many of these concepts would be difficult to understand by Deaf). Due to this aspect, experts advise to reduce the number of questions.
- Finally, the scale used: number of levels and the names for the different levels. For the number of levels, the expert panel advise 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. Forcing a user to decide provokes that this user asks more questions to the interpreter in order to understand all the details.

8 Conclusions

This paper has described a methodology for developing a speech into sign language translation system considering a user-centered strategy. This methodology consists of four main steps: analysis of technical and user requirements, data collection, technology adaptation to the new domain and, finally, evaluation of the system. In this paper we have explained how to apply this methodology in order to develop two translation systems in two specific domains: bus information and hotel reception.

It is possible to conclude that many aspects can be updated automatically from the generated parallel corpus. Except sign specifications and translation rules that are the two most demanding tasks. In the case of sign generation, it is useful a new editor for sign specification that incorporates new options for reducing significantly the sign specification time. In the case of translation rules, the authors are considering to work

over the statistical translation module in order to increase its performance and to delete the rule-based method.

9 Acknowledgments

The work leading to these results has received funding from the European Union under grant agreement n° 287678. It has also been supported by TIMPANO (TIN2011-28169-C05-03), INAPRA (MICINN, DPI2010-21247-C02-02), MA2VICMR (Comunidad Autónoma de Madrid, S2009/TIC-1542), CONSIGNOS (Plan Avanza Exp N°: TSI-020100-2010-489) and the European FEDER fund projects.

10 References

1. Brooke, J.: SUS: *A 'quick and dirty' usability scale*. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) Usability Evaluation in Industry, pp. 189–194. Taylor & Francis, London, 1996.
2. Cox, S.; Lincoln, M.; Nakisa, M.; Wells, M.; Tutt, M. & Abbott, S. The Development and Evaluation of a Speech-to-Sign Translation System to Assist Transactions International Journal of Human-Computer Interaction, 2003, 16, 141-161
3. Dreuw P., Ney H., Martinez G., Crasborn O., Piater J., Miguel Moya J., and Wheatley M., 2010. *The Sign-Speak Project - Bridging the Gap Between Signers and Speakers*. In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, 2010a. pp 73-80.
4. Dreuw P., Forster J., Gweth Y., Stein D., Ney H., Mar-tinez G., Verges Llahi J., Crasborn O., Ormel E., Du W., Hoyoux T., Piater J., Moya Lazaro JM, and Wheatley M. 2010. *SignSpeak - Understanding, Recognition, and Translation of Sign Languages*. In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010b. pp 65-73.
5. Efthimiou E., Fotinea S., Hanke T., Glauert J., Bowden R., Braffort A., Collet C., Maragos P., Goudenove F. 2010. *DICTA-SIGN: Sign Language Recognition, Generation and Modelling with application in Deaf Communication*. In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010. pp 80-84.
6. Elliott, R. and Glauert, J.R.W. and Kennaway, JR and Marshall, I. and Safar, E. *Linguistic modelling and language-processing technologies for Avatar-based sign language presentation*. Universal Access in the Information Society, Vol. 6, No. 4, pp 375-391, Springer, 2008.
7. Hanke T., König L., Wagner S., Matthes S. 2010. *DGS Corpus & Dicta-Sign: The Hamburg Studio Setup*. In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010. pp 106-110.
8. Huenerfauth, M.; Zhao, L.; Gu, E. & Allbeck, J. *Evaluation of American Sign Language Generation by Native ASL Signers* ACM Transactions on Accessible Computing, ACM, 2008, 1, 1-27
9. López-Ludeña, V., San-Segundo, R., Montero, J.M., Córdoba, R., Ferreiros, J., Pardo, J.M., *Automatic Categorization for Improving Spanish into Spanish Sign Language Ma-*

chine Translation. In press Computer Speech and Language, ISSN: 0885-2308, Available online 1 October 2011

10. Möllera, S., Smeeleb, P., Bolandc H., and Krebbera, J.,. *Evaluating spoken dialogue systems according to de-facto standards: A case study*. Computer Speech & Language. Volume 21, Issue 1, January 2007, Pages 26-53.
11. San-Segundo R., Barra R., Córdoba R., D'Haro L.F., Fernández F., Ferreiros J., Lucas J.M., Macías-Guarasa J., Montero J.M., Pardo J.M, 2008. *Speech to Sign Language translation system for Spanish*. Speech Communication, Vol 50. 2008. pp. 1009-1020.
12. San-Segundo, R., Montero, J.M., Córdoba, R., Sama, V., Fernández, F., D'Haro, L.F., López-Ludeña, V., Sánchez D., and García, A., 2011. *Design, development and field evaluation of a Spanish into sign language translation system*. Pattern Analysis and Applications. To appear.
13. Tryggvason, Judy., 2004. *VANESSA: A System for Council Information Centre Assistants to communicate using sign language"* School of Computing Science de la Universidad de East Anglia.