SPANISH SIGN LANGUAGE INTERPRETER FOR MEXICAN LINGUISTICS

Arturo Pérez

University ITS Chapala at Mexico
aperez@itschapala.com

Abstract

We present here the first visual interface for a Mexican Spanish Sign Language translator on its first development stage: sign-writing recognition. The software was developed for the unique characteristics of Mexican linguistics and was designed in order to use sentences or a sequence of signs in sign-writing system which are decoded by the program and converted into a series of images with movement that correspond to the Mexican sign language system. Using a lexical, syntactic and semantic algorithms plus free software such as APIss from Java, video converter software, data base manager like MySQL, Postgres and SQlite, was possible to read and interpret the rich and complex Mexican language. Our application for visual interface showed to be capable of reading and reconstruct each sentence used for the interpreter and translate it into a high definition video. The average time of video display vs number of sentences to interpret, probed to be in linear relation with an average time of two seconds per sentence. The software has overcome the problem of homonym words frequently used in Spanish language and verb tense relation for each sentence, special symbols such as #, %, $, etc. are still not recognized into the software.

Key words: Mexican Sign Language (MLS), spoken language translation, sign animation, sintactic algorithm

1 Introduction

The Sign Language is a system employed to stablish communication between persons with dis-capability both auditive and phonetic. A person with such dis-capability faces several obstacles while integrating into society. In order to overcome such difficulties (between a regular person and a deaf or deaf-mute person) Sign Language Interpreters softwares have been developed to attend this imperative need of communication. In the last two decades there has been more and more advances in visual interfaces for Sign Language Interpreters (Pardoa et al 2009, Halawani 2008, Dyng et al. 2008, Prada et al. 2008, Barra et al. 2007, Masakata 2006, Meurant 2004, Nyst 2004, Endbarg-Pederson 2003, Stouke 1960). The development of a Sign Language Interface

(SLI) strongly depends on the country where the Sign Language is used since not only the SLI is variable within different countries but also each country has its own characteristics for their official languages. It is well known that Mexico uses Spanish as its official language, however, the mexican linguistics (ML's) differ abruptly between the Spanish used in other countries such as Spain for example or the kind Spanish linguistics used in south American countries such as Colombia. This is why on the present date there are still many issues and misunderstandings towards persons who have a knoledge os Spanish different than the mexican linguistics and use it triyng to communicate in Mexico. Hence, translating Mexican text into Mexican Sign Language (MSL) requires a unique and special knowledge within this characteristic language. SLI's for Spanish language have been already produce for Rodriguez (1991), Prada et al.(2008) and Pardoa et al (2009), all of these works were produced only for European spanish linguistics and use 3D avatar technology to translate text into SL. To the date the are no works related to a Mexican Sign Language Interface (MSLI).

Producing a low cost software for Mexican language has been 47 a priority in Mexico ever since 2006, Mobil hardware devices have been created as SL Interpreter (Leybo´on et al. 2006), nevertheless, the software applied to this new mobil device was inefficient to decoding ML's, the system does not accounts for different hand positions, place of the hand gesture, hand direction and most important: face expression, which is one of the prevailing factors for a deaf-mute communication, since emotional expression plays a decisive factor on add meaning to each prhase for deaf-mutes (private communication Desarrollo Integral de la Familia, DIF, Jalisco).

This device (Leybo´on et al. 2006) was too robust and expensive for mass production. The consequences of not having such tools for deaf and deaf-mute persons has created an enormous incapacity of communication among the society added with discrimination factors towards the person with such discapability.

Another factor to be taken into account at the time of creating a new type of SLI, points at the unique cultural characteristics present in each country and then, with different priority of basic needs. The tools developed for our MSLI originate on the basis of two main priorities for mexican society: (a) the need to communicate from one person to another in order to obtain and provide one or more services and (b) the need of the person with a dis-capability to have a meaningful response to this communication. On the latter issue, mexican persons with discapability find it more meaningful to interact with a video image that can show up a sensible emotion than a 3D avatar.

While 3D avatars can indeed be constructed to represent an emotion while displaying, it has been prooved for local experts on SL (Desarrollo Integral para la Familia, Jalisco, private communication) that Mexicans do not relate well with virtual images while trying to communicate a feeling or necessity.

At the same time we do not discriminate the advantage of having 3D avatars for the case of simple written instructions or web page translation in a human-machine or in a more general perspective: human-object interpretation, here, Prada et al. (2008) offers the best deal for the interaction when communication 74 between people it is not a prime manner or can be avoided.

As an overall picture, the development of the MSLI does not involves any new relevant work in the rea of signal processing, the algorithms used in this work remain the same known at date, however, this is a work that focuses on the development area of engenieering, meaning a new practical and inovative tool devoloped to meet current social needs in the mexican society. The outline of this paper is presented as follows: section 2 describes the material and methodology employed inorder to explain the differences of an SLI for mexican language.

2 Methodology

In order to have an appropriate translation from speech transcriptions into SL it is necessary to have a parallel corpus institution to fit the translation models, test them, evaluate them and have them corrected in each phase of the process. In our case, the development of our MSLI was performed with the aid of a well founded federal institution: Desarrollo Integral para la Familia de Jalisco (DIF, Jalisco), which is a solid foundation institution in Mexico dedicated to aid persons with this type of discapability (among other functions in its primary activities). DIF, Jalisco, provided us with the unique opportunity to work with several experts (making a total group of 8 instructors for MSL) in order to test and apply the MSLInwith their respective students. This advantage gave us the opportunity to attend to the basic and real needs for a deaf or deaf-mute person in our society.

At the time of working with this corpus it was noticed that a 3D avatar would not be helpful at adressing person-person translations, where, a 3D avatar is more suitable to adress a human-object translations in a good feasible way. A mayor emphasis was pointed from DIF Jalisco over to a deaf-mute person having a real meaningful communication, hence, the need of a consistent translating system involving human facial expression. In this manner, itwas choosen the use of interactive videos as the apropiate way 99 of providing such answer.

Table 1. Software and Hardware

Employed software	Employed Hardware	Sampled Group
APIss (Java)	video camera	MSL students (7 children)
video converter software		MSL teachers (3 adults)
data base manager MySQL	keyword	
Postgres	screen	
SQlite		

The development of the platform was divided into four main stages: character recognition modulus, the syntactic modulus, the semantic modulus and the syntax modulus, these stages explain a sufficient coordinated method and a good efficiency degree of quality.

Figure 1. From left to right: (a) SLI Man window. (b) Interpreter screen, (c) Introduced sentence for SLI. The window of the SLI contains a text field were the sentence is introduced, this sentence must be written in simple tenses, the system accepts lower and upper case text.

The first stage (Figures 1,2) is the character recognition process performed with a lexical algorithm. The design of our lexical algorithm consist on a double buffer compiler system. For the double buffer system we have taken the couple used by Aho et al.1990.

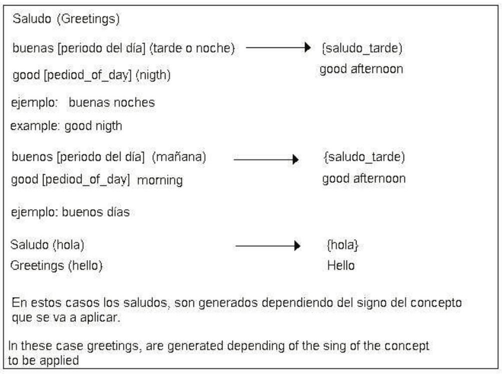

Figure 2. First type of simple sentence, greeting.

The lexical analysis consists on the identification of the word or sentence and the elimination of not usable words for the MSL. Within this process the text to be translated is first captured on screen using a basic keyboard. Afterwards, the system indicates to the user if any invalid character has been typed, the invalid characters in this stage are #, %,¯ , $, ,´ .¨ , etc. Once all characters are read , the process of translation is allowed to continue only if all characters are recognized as valid characters. If the program finds one invalid character the translation process stops and a warning message is displayed on screen asking for the text to be retyped.

The second stage consists on the syntactic analysis. The algorithm applied was the normal form of Chomsky's algorithm (Chomsky 1965) this process consists on the identification of the sentence structure: subject, predicate, nouns, conjunctions, verbs, transitive verbs, intransitive verbs, complements, adverbials, and the use of the tences on each verb in the sentence (present, past, and future tences). During this process words like articles or conjuntions in the sentence are eliminated (since sign language does not uses any of those). Once the structure of the sentence is analyzed the process is allowed to continue (Figures 3,4).

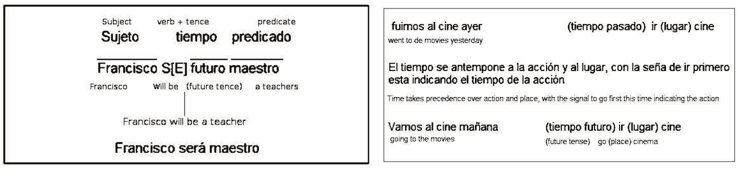

Figure 3. Construction of a sentence using simple tenses.

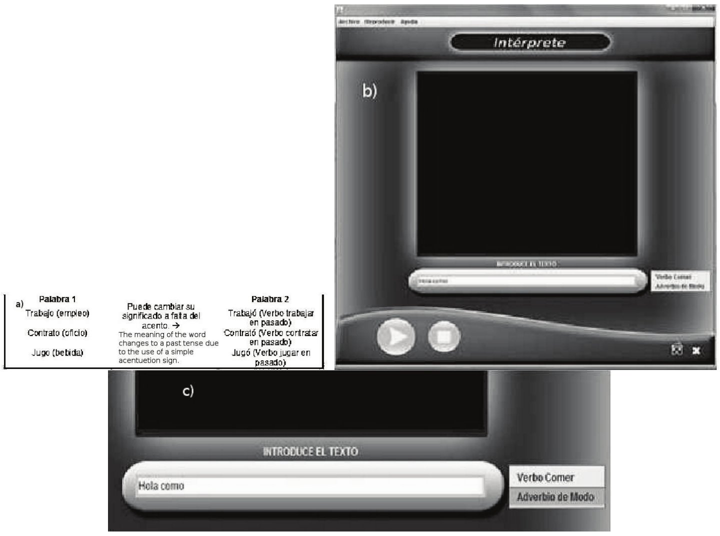

Figure 4. From left to right: (a) Change of meaning in the word according to accentuation sign. (b) current banner for unknown sign, (c) Sentence selection in the case of homonym and homograph word. In (b) the SLI system is programed to detect homonym words on the database by showing an current banner. The desired meaning for the sentence or word is given to be chosen afterwards.

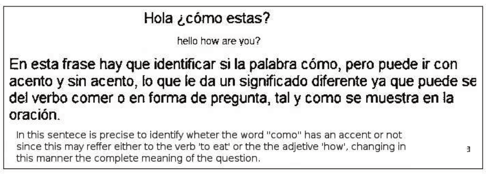

Figure 5. Changes in the meaning of a sentence depending on grammatical sign applied in the sentence.

Figure 6. Form left to right: (a) Error window. (b) Correction Screen, (c) Sentence. The SLI system does not adtmits signs such as $,%,&, dot, colon. If the user types an invalid character the systems displays a message on (a). The system can is able to interpret no more than 12 digits and the structure of this should be in a continuous form (without separations or blank spaces).

On the third stage we find the semantics analysis, this process uses the algorithm to arrange the main sense of the sentence and eliminate any ambiguous senses. The main objective of this process is to identify either homonym and homograph words (words that sound or are writen in a similar way such as casa or caza, both existent on the mexican language and with different meanings depending on the form they are aplied in each sentence). This stage was completed at a 70% rate due to the complexity of the algorithm, in order to complete this stage an emergent banner appears at the time the user uses an homograph or homonime word, the banner shows up a menu with options for different meanings and gives the user the option to choose the more convenient one.

The last stage of the SLI is the generation of the sequence of images selected from a re spective database of the sign language. Once the semantics is produced, the syntax algorithm(tal de tal ao) simply uses the before structure to select and display the appropriate video providing then the expected translation.

3 Results

Our main results are described as presented in the methodology. In the first stage we obtained the products of sentence typing as well as the voice recognition process, it was mesured that this process took an average time of 2.5 seconds per recognized sentence. The average time of pattern recognition is faster when the sentence is only written instead of spoken, this caveat waves on the fact that voice capture processes are still not fully undestood and keeped yet under development.

For the second stage of the MSLI it was possible to identify subject, predicate and verb within the sentence's grammar and distinguish between the present, past or future tences for the verb (since most of SL does not uses verb tences) at this point it was also possible to eliminate other words not used by MSL such as conjunctions or intransitive verbs. As shown in Figure 4 the recognition process of the SLI produces a screen for the video sequence and a space at the bottom to show the written or spoken sentence. This typed sentence either, follows or it is made to adapt to the grammatical order of Mexican linguistics (noun, verb, predicate, et) in order to continue the translation. The process aldo stops when a unknown symbol is typed or a unrecognized word is spoken (such as caaa instead of casa).

Afterwards, the MSLI produces the correspondent video in the database and differentiates similar sentences with similar words but different gramatical puntuation (such as the spanish mexican accent, wich then adds a total different meaning to the sentence).

The semantic modulus was completed at a 70% due to the complexity of the algorithm dealing with homograph words. In order to solve complete such inconvenience a recursive help banner was applied, every time the user types an homograph word a graphic menu appears with several options to choose from. Once the convenient option is chosen by the user the interpretation process continues and the video is displayed. For the case of compound sentences the SLI produces only a single video to show with the main ideas of the compound sentence.

Within our results it was also created a small buffer during the process of lexical analysis. This buffer is a thread manager created to storage words which might be captured with voice recognition in the implementation of future hardware device for sound recognition, not yet employed at this stage of the SLI but left aside in order to focus only on this first stage of SLI, written sentence recognition transformed into video image. Finally it was noted that the quality of the final videos shows a small flash in between videos this caveat is due to the multimedia java version employed for the development of MSLI.

4 Discussion

The obtained results with the SLI were taken to DIF Jalisco and it was proved how the general needs demanded at time were mostly covered. Regarding the employed algorithms used in this SLI: lexical, syntactic and semantic, it is plausible to modify and generate the current syntactic and semantic algorithms employed here as mentioned by Montero (2004) and Earley (2001). As for the lexical algorithm it is now currently used the standard algorithm of Tokens since it only consists on pattern recognition and there are not new and specifics needs that reacquires to employ or perform any kind of modification to it.

An important fact to be addressed is our use of video images instead of a 3D avatar such as the case of Prada et al. 2008 or Halawani 2008. This difference creates both advantages and disadvantages regarding to the area of application for the SLI, the fact of having a 3D avatar results very convenient at the time of simple interaction for instructions or guidelines using tools such as web pages or any other human-machine interface, nevertheless the 3D avatar faces limitations at the moment of human-human interactions where the occurrence of of facial gestures takes a mayor role to be taken into account in order add meaning to the conversation for a person with this type of discapability. Adding more detailed expression to this caveat has a high cost in development but if such improvement could be achieved this type of tool on the SLI could have a mayor impact on many areas of interaction for a deaf-mute person.

The fact of a deaf-mute interacting with another person creates 186 a basic need to receive some type of facial gesture in each sentence in order to have a complete meaning of the conversation. In this case, the turn point comes into crating a complex conversation, hence, having a complete sequence of images to follow up such type of interaction, the use of video images faces such difficulty and it is been left as future work.

During the process of translating a sentence into image the average time resultant in each process (2 seconds) clearly indicates that each algorithm and the employed process produces a reliable and consistent response that connects with sufficient feasibility the continuity of the translation process as such. The relation of the main variables with time (Figure 1) and between, presents a linear correspondence. This linear correspondence is the mathematical proof of how each one of this variables has an independent role in the process besides that the orthogonality between them it is also a way to describe the evolution of the system.

As for the flash seen between images for translated sentence it was noted that such effect was due to operative system and its different versions. This flaw is strongly seen for Windows Operative System (OS), specially, Windows Vista and Windows 7. The effect greatly diminishes at Windows XP, while for Mac OS there is no appearance of such flash. We strongly believe that such discontinuity is due to the Java Media Framework API for Windows which most probably needs a new compatibility upgrade for the multimedia version for Windows, which is not the case of the Mac OS.

The case of a SLI as a new tool for Mexican linguistics has the advantage of being the first tool in this country capable of reproduce and translate simple human-human conversations whereas it has been mention the appearance of an photo-electric sensor for hand movement as a SLI tool (Leybon-Ibarra et al. 2006) this particular software uses a photo-electric electrode system within an adaptable glove with 3D avatar screen images to produce the translation, nevertheless, the main limitation of such system consists in the incapability of freedom of hand movement, this is: on hand direction, orientation, crossed or bended finger positions and facial expression.

5 Conclusions

The new SLI for Mexican language was presented in this article, our main results showed to be reliable and satisfactory within the selected proof sample which were selected within the current needs of the actual MSL for Mexicans and specially aimed at the expressed needs of DIF Jalisco. The program showed to be capable of overcoming many MSL difficulties such as homograph and homonym words to construct a sentence.

Our visual interface showed to be capable of reading and reconstruct each sentence used for the interpreter and translate it into a high quality video. The average time of video display vs number of sentences to interpret, took an average time of 2 seconds per sentence and probed to behold within a linear relation, which shows how both variable are independent between them (hence able to describe different characteristics of the evolution of the system).

Nevertheless, some caveats are still to be considered and investigated such as the use of special symbols such as #, %, $, etc. which are still not recognized into the software. The use of complex compound sentences are still yet not recognized by the SLI and has to be taken into account for more elaborated dialogues. More work for facial gesture recognition has to be done since most of SL persons uses them to change the meaning of each sentence and it is also a way to recognize the meaning of a phase used to communicate with them. In the mean time, the use of SLI for an average conversation and average needs of a deaf-mute person with a regular non-sapient MSL person is now covered and capable to perform its main task.

References

1. Aranda., B. E., 2008. La vulneracin de los derechos humanos de las personas Sordas en Mxico. Comision Nacional de los Derechos Humanos, CNDH.
2. R. Barra, R. Crdoba, L.F. Haroa, F. Fernndeza, J. Ferreirosa, J.M. Lucasa, J. Macas-Guarasab, J.M. Monteroa and J.M. Pardoa, *Speech to sign language translation system for Spanish*, Aplied Soft Computing, 2008.
3. Comparn, J. J., 1999. Lengua Espaola I. Mxico: AMATE. Discapacidades, E. C. *La sor dera y la prdida de la capacidad auditiva.,* http://www.sitiodesordos.com.ar/sordera.htm.
4. Ding Lilia, Modelling and recognition of the linguistic components in American Sign Language, Aplied Soft Computing, 2008, 421, 105.
5. Dons, R., & Ortz, C., 2005, XXXV *Simposio Internacional De La Sel*: http://www3.unileon.es/dp/dfh/SEL/actas.htm.
6. J. Earley, *An Efficient Context-Free Parsing Algorithm, PhD tesis,* University of California, Berkeley, California, 1970, pp. 94-102, http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/cmt-55/lti/Courses/711/Class-notes/p94-earley.pdf.
7. El Universal. 2006, : http://www.eluniversal.com.mx/articulos/30484.html.
8. Estrada, B., 2008, Sordos: www.sordos.org.mx/articulo.doc.
9. Galicia, S. N., 2000, Instituto Politcnico Nacional Centro de Investigacin en Computacin Laboratorio de Lenguaje Natural,Anlisis sintctico: http://www.gelbukh.com/Tesis/Sofia/tesisfinal.htm.
10. Garca, J. R., *& Giner*, B., 2007, Pearson, Prentice Hall.
11. Leybon I. J, Ramirez B. M.R., Picazo T. V., Photo-Electric Sensor Applied to Hand Fingers Movement, Computacin y Sistemas, 2006, 10, 556.
12. Lodares, J. R., *Aplicaciones Lexemticas a la Enseanza Del Espaol*, 2009, Clarn, Revista de Nueva Literatura, 78.
13. J. M. Montero M., *Desarrollo de un Entorno para el Anlisis Sintctico de una Lengua Natural,* Universidad Politcnica de Madrid, Espaa, 2004, http://lorien.die.upm.es/juancho/pfcs/JMMM/pfcjmmm.pdf.
14. Nuno, R. (1998). Correlatos neurofisiolgicos del lenguaje de senas en el nino sordo. Proyeto de Investigacin.
15. J.M. Pardoa, J. Ferreirosa, V. Samaa, R. Barra-Chicotea, J.M. Lucasa, D. *Sn-chezb and A. Garcab Spoken Spanish generation from sign language*, 2009, Aplied Soft Computing, 123.
16. Dr.Sami M.Halawani, *Arabic Sign Language Translation System On Mobile Devices*, International Journal of Computer Science and Network Security, 2008, 8, 1.
17. Suphattharachai Chomphan, Towards the Development of Speaker-Dependent and Speaker-Independent Hidden Markov Model-Based Thai Speech Synthesis, 2009, Journal of Computer Science, 5, 905.