Universal Access in Human-Computer Interaction. Applications and Services: 6th International Conference, UAHCI 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part IV. Lecture Notes in Computer Science, Volumen 6768. Springer, 2011. 567-576.

DOI: http://dx.doi.org/10.1007/978-3-642-21657-2_61

Integration of a Spanish-to-LSE[1] Machine Translation System into an e-learning Platform

Fernando López-Colino, Javier Tejedor, Jordi Porta and José Colás

[1] Human Computer Technology Laboratory, Escuela Politécnica Superior,
Francisco Tomás y Valiente 11, 28049 Madrid, Spain
{fj.lopez, javier.tejedor, jordi.porta, jose.colas}@uam.es

Abstract. This paper presents the first results of the integration of a Spanish-to-LSE Machine Translation (MT) system into an e-learning platform. Most e-learning platforms provide speech-based contents, which makes them inaccessible to the Deaf. To solve this issue, we have developed a MT system that translates Spanish speech-based contents into LSE.
To test our MT system, we have integrated it into an e-learning tool. The e-learning tool sends the audio to our platform. The platform sends back the subtitles and a video stream with the signed translation to the e-learning tool.
Preliminary results, evaluating the sign language synthesis module, show an isolated sign recognition accuracy of 97%. The sentence recognition accuracy was of 93%.
Keywords: Machine Translation, Sign Language, LSE, Accessibility

1 Introduction

During the last decade there has been an increasing interest in machine translation research from an oral language to a signed language. To obtain a complete system providing this functionality, it is necessary to integrate the results from three different research areas: Automatic Speech Recognition (ASR), Machine Translation (MT) and Sign Language (SL) Synthesis. Although the literature provides several examples of working prototypes, it is hard to find examples of these systems applied to education with representative results (see next section). Different statistics report that the 47% of the Spanish Deaf people have not finished any degree or even are illiterate, compared to the 21% of the whole Spanish population (INE[2] 1999 and MEC[3] 2000/2001). Such data claim that Deaf people have many drawbacks to achieve the educational, social and cultural levels of the general society. It is possible to use MT systems to contribute to alleviate this social problem.

---

[1] Spanish Sign Language
[2] National Statistics Institute
[3] Ministry of Education

2 Related Work

We have stated that a system that converts speech into a signed message requires using three different research areas: ASR, MT and SL synthesis. We will briefly describe recent works from the last two. Finally we will comment some MT translation systems that have been applied to the e-learning.

In 2003, a technical report about MT systems applied to ASL [1] concluded that existing systems were just work in process or simple demonstrators. In the recent years, different authors have reported improvements to their systems [2, 3] and applications to other languages [4, 5]. However, we believe this research direction has not reached the level of development achieved in other areas like MT between oral languages. These systems do not fulfill the required functionality to provide a real useful service to the Deaf community.

Although the SL-related MT systems do not provide complete functionality, the literature provides some examples of applications that enhance e-learning systems by means of synthetic signed contents. Some systems use the existing technology to provide e-learning tools to learn a SL [6, 7]. Although these systems only allow the visualization of signed contents, there are others that also allow the interaction of the student and provide feedback about his/her signing ability [8]. Particularly, for Spanish Sign Language (LSE), a tool was recently presented that aims at helping the study of new vocabulary [9]. However, our aim is to improve the experience of deaf users when using e-learning tools. The application of previous MT systems to e-learning tools allows deaf users the access to the speech-based contents in these applications. Some of these systems have only focused in a closed domain, such as mathematics [10, 11]. Finally, it is remarkable the efforts of the Greek researchers for generating an e-learning platform for Greek Sign Language [12, 13, 14].

3 System Overview

ISABEL [15] is a multipurpose video streaming platform, which allows easy modification of the viewer's layout and the management of video streams. We have used the ISABEL system due to its e-learning capabilities. The speakers' audio stream is both transmitted to every student and to our MT system (named MaTSyLSE, depicted in Fig. 1). The platform receives the audio stream and returns to the ISABEL platform both the transcribed text, used for subtitling, and the video stream of the signed translation. The ISABEL platform considers this signed video stream as another participant in the conference management and redirects the stream to the users that request for it. Fig. 1 shows both the architecture of the system and the final GUI of the working system.

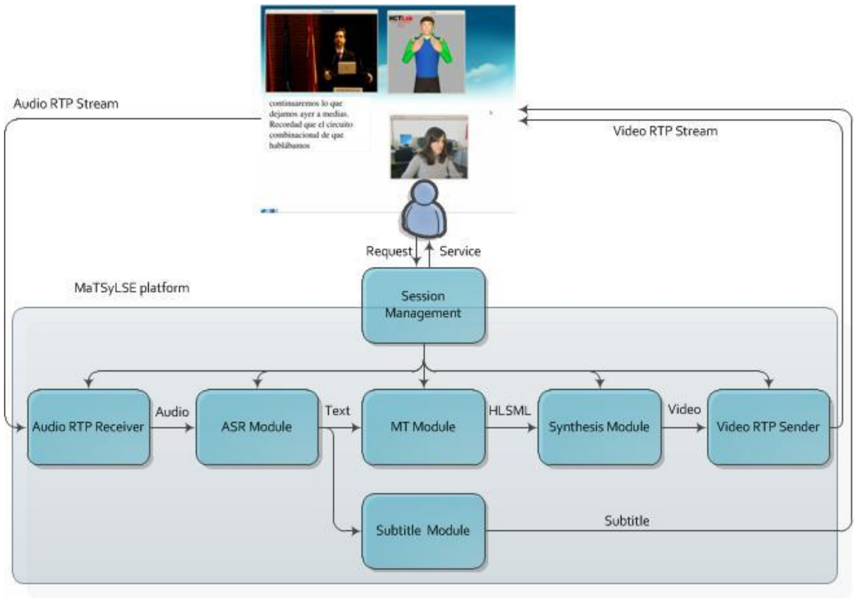

Fig. 1. GUI and diagram of the architecture.

3.1 ASR Module

The audio obtained from the video stream is transcribed. In so doing, the ASR module consists of a speech recognizer built from Julius tool [16]. It is a large vocabulary continuous speech recognition system that works on real time to produce the most likely sequence of words corresponding to the input speech signal. It uses Hidden Markov Models to represent the acoustic space, a bigram language model and contains a vocabulary of 5K words. The output of this module is the transcription of the speech.

3.2 MT Module

When translating between similar and intelligible languages, as occurs between many Romance languages, quality and massive translation is possible to some extent. However, translation between typologically unrelated languages, as Spanish and LSE are, has another purpose: to break the so-called language barrier that makes communication between individuals difficult, if not impossible.

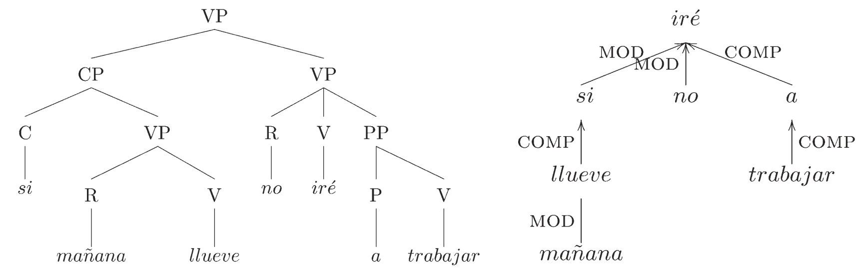

Fig. 2. Constituency tree (left) and its corresponding dependency tree (right).

The MT module implements a transfer-based approach to translation. The main phases in this architecture are analysis, transfer and generation. In the analysis phase, a Spanish text like *Si llueve mañana no iré a trabajar* ("if it rains tomorrow I won't go to work") is first analyzed morphologically and then syntactically to obtain a constituency tree like the tree on the left side of Fig. 2. Then a dependency structure, depicted in the right side of Fig. 2, is built up from the constituency structure as an intermediary result for transfer. A dependency structure is a more abstract structure where constituency and word order have been removed and only functional relations between words are represented. This structure is transferred to LSE. During the process, some nodes representing words are removed, as is the case of definite articles, which have no correlate in LSE. On the other hand, other information can appear in the new structure, like the non-overt subjects in Spanish. Note that in the transferred dependency structure of Fig. 3, a first person pronoun subject *yo* ("I") has been made explicit. Generation is formulated as a constraint problem. A precedence graph is built from the information about the relative order between a head and its dependent or between pairs of dependents with the same head. As an example: suppose a noun phrase with the structure "N_1 *de* N_2" (N_1 of N_2). When translating this expression to LSE, sometimes N_1 precedes N_2, as in *hermano de Luis* ("Luis' brother"), which is translated as LUIS HERMANO because *hermano* denotes a permanent relationship. In other cases the correct order is N_2 N_1, as in *mesa de estudio* ("study table"), because *estudio* is the function or purpose of *mesa*, and therefore, it is translated as MESA ESTUDIO. An algorithm based on topological sort produces the linear surface order in LSE. Consequently, the dependency graph of Fig. 3 will generate the LSE glossed sentence EJEMPLO / MAÑANA LLOVER / YO TRABAJAR IR NO ("EXAMPLE / TOMORROW RAIN / I TO-WORK TO-GO NO"). Finally, as a last step before the synthesis, signs are morphologically realized according to their description.

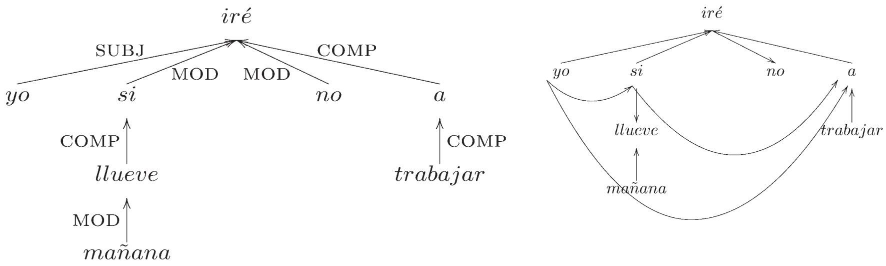

Fig. 3. Transferred dependency tree (left) and its corresponding precedency graph (right).

A wide-coverage unification-based grammar for Spanish has been implemented for the analysis phase. The linguistic phenomena dealt by the grammar include complementation, adjunction, pronominalization, relativization, etc. Grammar is inspired by the lexicalist approach of theories as the Head-driven Phrase Structure Grammar (HPSG) [17], where the lexicon is the locus of almost all the grammatical information and schemas can be implemented as context-free rules. However, for convenience, the accounting for complex groups of specifiers, modification or coordination is not expressed in the lexicon but in rules modeled under X-bar theory [18]. Apart from schemas, the grammar also includes lexical rules. A lexical rule is an input-output rule mainly used, in our case, to alter the argument structure of a lexical item. Therefore, phenomena like subject pro-drop, complement extraction, complement deletion, animated object marking or pronominalization, to name a few, are dealt with lexical rules. The parser is an active chart-based agenda-driven Earley's parser [19] implemented in Prolog but makes use of the language processing services implemented in FreeLing [20] for tokenization, named-entity recognition and PoS-tagging.

Parsing robustness is achieved through three techniques: relaxation, weighted phrase structure rules and a mechanism for the assembly of partial parses. Relaxation here is used mainly to handle frequently seen incorrect assignments of the PoS-tagger by means of casting rules. With these rules, for example, determiners and pronouns recover their original lexical determiner-pronoun ambiguity, so that the parser can proceed exploring, if necessary, a wider search space for a complete analysis. Semiring parsing [21] is a framework in which deductive systems for parsing can be extended to manipulate weights. Semirings generalize the representation and computation of weights for weighted languages. Examples of semirings are the counting semiring, defined as $(\mathbb{N}, +, \times, 0, 1)$, which is used to compute the number of derivations in grammars with unit rules, or the Viterbi semiring $([0..1], \max, \times, 0, 1)$, which is used to compute the probability of the best derivation for a probabilistic grammar. Here, we use the tropical semiring defined as $(\mathbb{N}, \min, +, -\infty, 0)$ to compute the minimum derivation weight of a parse tree. This decision pretends, as in other preference-based parsers, to favor structurally simple analyses. Speakers make mistakes, correct themselves during speaking, produce false starts and use ungrammatical constructions. Several methods were proposed for robust analysis. Partial parse is defined in [22] as the set of intermediate passive edges of the chart whose spans are not overlapping and cover the entire input sequence. For selecting the best sequence of partial parses, they apply the shortest-path algorithm to maximal projections, i.e., complete phrases. These phrasal edges are assigned a weight of one whilst the weight of lexical edges is two. Finally, solutions are re-ranked using a scoring function. Shortest-path algorithm is an adaptation of the general algorithm to the case of direct acyclic graphs. The algorithm runs with a worst-case time linear in the number of edges and vertices. In [23], different methods for partial parse selection are presented and evaluated for HPSG parsers producing much deeper analyses than the parser presented here. The criteria examined for selecting partial parsers include longest edges and shortest paths using several weight estimation functions based on probabilities derived from treebanks. Our heuristic approach here is simple: it uses the derivation cost and is based on the intuition that very complex partial analyses should be rejected as well as the simplest ones. So, arcs [*i*,*j*,*w*] in the chart are assigned a distance *w*/*K(j-i)* where *K* is a factor that controls the complexity of the analyses and can be estimated from the weights and lengths of partial analyses within complete analyzed utterances.

3.3 Synthesis Module

The synthesis module [24, 25] receives the translated message from the MT module using the HLSML notation. This notation allows representing glosses, phonetic definitions, morphologic variations, classifier predicates, fingerspelling sequences, etc. Although the synthesizer can manage all of these inputs, the first version of the MaT-SyLSE platform is based only in glosses.

The system parses the HLSML input by obtaining the sequence of signs in the message. The phonetic definition of each sign is stored in a relational database. This definition is used for the generation of the animation sequence of the whole sentence. This animation is rendered using a real time rendering API [26] and the resulting video is streamed to the client.

3.4 Subtitle Module

The subtitle module is a simple text processing element. The ASR module generates different length transcriptions, from one word to a complete sentence. These transcriptions cannot be directly redirected to the users because they may collapse the subtitle window, making its reading impossible.

The subtitle module buffers the transcriptions and sends them to the user controlling the length and the pauses between consecutive text chunks. These chunks' average length is 18 characters sent every 700 ms; the final length is adjusted so no word is hyphenated.

4 Evaluation

The initial evaluations of the system have been focused on the evaluation of the user GUI required for deaf users and the avatar's signing performance. Evaluation of the ASR and the MT modules will be presented in a forthcoming work.

For the evaluations, a group of ten deaf native signers have been selected; four of these users work at the LSE linguistic department of the FCNSE. The ages of this group ranges from 20 to 62 years old, with an average of 37.1 years. All of the users in this group have a medium comprehension level for reading Spanish. The evaluations contained written instructions that users could understand.

4.1 Definition of the GUI

The first evaluation was oriented to the definition of the layout of ISABEL for this kind of application. We proposed a set of five different window distributions with different sizes. The unanimous selection was the set with the largest avatar window size. This is understandable because the signer users will omit the speaker's window and they prefer as much detail as possible in the avatar's window.

The preferred layout is depicted in Fig. 4. The final avatar dimension is 600×600 pixels. Although signing users preferred a larger avatar size, a compromise between the screen size and the transmission bandwidth was set. There is also a subtitle window for non-signer users. The font of the window can be adjusted by the user.

Fig.4. Screenshot of the application.

4.2 Evaluation of the synthesis

Preliminary evaluations have measured the performance of the Synthesis module. These evaluations measured both the isolated sign recognition rates and the sentence understanding. The previously described group of native LSE signers performed both the evaluations.

We provided the users with the different URL to access the different web forms, which contained the questionnaires. These forms contained text that explained the experiment to the users and instructed them how to proceed.

The first questionnaire contained a set of twenty randomly-selected signs. Each question showed a video of the avatar signing each sign. After the visualization of each video, the users had to write the Spanish word that corresponded to that sign. For some of the selected signs there are multiple correct answers (e.g., LSE's sign for the noun "water" and the verb "to drink" is the same). We considered as correct all of them. The result of this evaluation presents a rate of 97% correctly recognized signs.

The second questionnaire contained a set of ten sentences in LSE. These sentences were proposed by an LSE interpreter and synthesized using the synthesis module. Hence, we ensured that these sentences were grammatically correct. The procedure of this second evaluation was similar to the one stated above. After the visualization of the sentences, the users had to choose one out of five different possible sentences. These sentences were different variations to the one represented by the avatar. The result of this second evaluation shows a 93% of sentences correctly recognized.

After the evaluations, we asked the users for their opinion. One of their concerns was the relatively inexpressiveness of the avatar, because the facial expression was only modified due to phonology and syntax requirements. Another concern was related to the intelligibility of some of the signs used in the sentences. Some users misunderstood few signs and selected a wrong answer.

5 Conclusion

This paper has presented a Spanish-to-LSE machine translation system integrated within ISABEL, an e-learning platform. On one hand, the system allows deaf people to access the speech-based information by means of a synthetic signed message using MT techniques. On the other hand, the system also provides subtitles for people with low degree of hearing loss.

As far as experimentation is concerned, this work has presented an initial evaluation of the system. We have focused on evaluating the visual layout of the accessible version of the e-learning tool. We have also provided initial evaluations of our SL synthesis module.

As future work, we will extend these evaluations, measuring the performance of the ASR and MT modules. We will also extend our research towards different educational domains (math, science, etc.) and towards different age groups (university, high school, primary).

The ISABEL platform is a general purpose videoconference tool. We have proposed using it for e-learning. However, we will also explore other environments such as VoIP online MT translation, broadcast news MT translation, etc. which can get benefit from our MT system.

Acknowledgements. Authors would like to acknowledge the FPU-UAM grant program for its financial support. Authors are grateful to the FCNSE linguistic department for sharing their knowledge in LSE and performing the evaluations. Many thanks go to María Chulvi and Benjamín Nogal for providing help during the implementation of this system. This work was partially supported by the Telefónica Móviles España S.A. project number 10-047158-TE-Ed-01-1.

References

1. Huenerfauth, M.P.: A Survey and Critique of American Sign Language Natural Language Generation and Machine Translation systems. Tech. Rep. MS-CIS-03-32. Computer and Information Science, University of Pennsylvania (2003)
2. Jennings, V., Elliot, R., Kennaway, R., Glauert, J.: Requirements for a Signing Avatar. In: Proceedings of LREC, 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. pp. 133–136.Valletta, Malta (2010)
3. Kennaway, R., Glauert, J., Zwitserlood, I.: Providing signed content on the internet by synthesized animation. ACM Transactions on Computer-Human Interaction 14(3): 1–29 (2007)
4. Dasgupta, T., Basu, A.: Prototype machine translation system from text-to-indian sign language. In: Proceedings of the 13th International Conference on Intelligent User Interfaces (IUI-08) pp. 313–316. ACM, New York, NY, USA (2008)
5. 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.: Speech to sign language translation system for Spanish. Speech Communication 50(11-12):1009–1020 (2008)
6. Agravat, B.B.: Inputted text to animated sign language, interactive interface, a self-learning with fun. In: ACM SIGGRAPH educators program. ACM, NewYork, NY, USA (2007)
7. Ohene-Djan, J., Naqvi, S.: An Adaptive WWW-based System to Teach British Sign Language. In: Fifth IEEE International Conference on Advanced Learning Technologies. pp. 127–129 (2005)
8. Aran, O., Ari, I., Akarun, L., Sankur, B., Benoit, A., Caplier, A., Campr, P., Carrillo, A., Fanard, F.X.: SignTutor: An Interactive System for Sign Language Tutoring. IEEE Multimedia 16(1): 81–93 (2009)
9. López-Colino, F., Colás, J.: New interactive tool for teaching new vocabulary in LSE. In: Proceedings of III National Conference on Spanish Sign Language. CNSE-UNED, Madrid, Spain (2009)
10. Adamo-Villani, N., Doublestein, J., Martin, Z.: The Mathsigner: An Interactive Learning Tool for American Sign Language K-3 Mathematics. In: International Conference on Information Visualisation. pp. 713–716 (2004)
11. Adamo-Villani, N., Carpenter, E., Arns, L.: An immersive virtual environment for learning sign language mathematics. In: ACM SIGGRAPH Educators program. ACM, New York, NY, USA (2006)
12. Efthimiou, E., Sapountzaki, G., Karpouzis, K., Fotinea, S.E.: Developing an e-Learning Platform for the Greek Sign Language, LNCS vol. 3118. K. Miesenberger, J. Klaus, W. Zagler (Eds.), Computer Helping People with Special Needs (2004)
13. Karpouzis, K., Caridakis, G., Fotinea, S.E., Efthimiou, E.: Educational resources and implementation of a Greek sign language synthesis architecture. Computers & Education 49(1): 54–74 (2007)
14. Sapountzaki, G., Efthimiou, E., Karpouzis, K., Kourbetis, V.: Open-ended resources in greek sign language: Development of an e-learning platform. In: Proceedings of the workshop on the representation and processing of sign languages. LREC-04. Lisbon (2004)

15. Quemada, J., de Miguel, T., Pavón, S., Huecas, G., Robles, T., Salvachúa, J., Acosta Ortiz, D., Sirvent, V., Escribano, F., Sedano, J.: Isabel: An Application for real time Collaboration with a Fexible Floor Control. In: Proceedings of the 1st International Conference on Collaborative Computing: Networking, Applications and Worksharing. San Jose, CA, USA (2005)
16. Lee, A., Kawahara, T., Shikano, K.: Julius – an Open Source Real-Time Large Vocabulary Recognition Engine. In: Proceedings of Eurospeech, pp. 1691–1694 (2001).
17. Pollard, C., Sag I.A.: Head-Driven Phrase Structure Grammar. Studies in Contemporary Linguistics. University of Chicago Press. Chicago (1994)
18. Jackendoff, R.: X-bar Syntax: A Study of Phrase Structure. MIT Press, Cambridge, MA, USA (1977)
19. Earley, J.: An efficient context-free parsing algorithm. Communications of the ACM 13(2), 94–102 (1970)
20. Atserias, J., Casas, B., Comelles, E., González, M., Padró, Ll., Padró, M.: FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. In: Proc. of LREC. (2006)
21. Goodman J.: Semiring Parsing. Computational Linguistics, 25(4):573–605 (1999)
22. Zhang, Y., Kordoni, V., Fitzgerald, E.: Partial Parse Selection for Robust Deep Processing. In: Proceedings of the Workshop on Deep Linguistic Processing (DeepLP-07), pp. 128–135. Prague, Czech Republic (2007)
23. Kasper, W., Kiefer, B., Krieger, H.U., Rupp, C.J., Worm, K.L.: Charting the Depths of Robust Speech Parsing. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics. pp. 405–412. ACL, College Park, Maryland, USA (1999)
24. López-Colino, F., Colás, J.: Hybrid paradigm for Spanish Sign Language synthesis. Universal Access in the Information Society (In press)
25. López-Colino, F., Porta, J., Colás, J.: Linguistic Principles Applied to the Automatic Synthesis of Spanish Sign Language. Estudios de Lingüística de la Universidad de Alicante 24 (2010)
26. Java Community Process: JSR-184. Mobile 3D Graphics API for J2ME™