Speech to Mexican Sign Language for Learning with an Avatar

Fernando Barrera Melchor[✉], Julio César Alcibar Palacios,
Obdulia Pichardo-Lagunas, and Bella Martinez-Seis

UPIITA-IPN, Insituto Politécnico Nacional, Mexico City, Mexico
barreramelchorf@gmail.com, jcesarap1@gmail.com,
{opichardola,bcmartinez}@ipn.mx

Abstract. This work proposes the combination of natural language processing techniques and the use of graphic development engines to create an avatar to display information translated into Mexican Sign Language (MSL). It describes the application of the guarantee analysis carried out on Mexican Sign Language in the implementation of a rules-based automatic translation system. Taking into account the graphic character of sign language, the generated translation is represented through the use of an avatar developed in a virtual environment Unity. The translation is based on a restricted vocabulary associated with the learning of history of Mexico for the fourth level of primary education. The grammatical structures most used in the history book were also analyzed and included in the rules system. The application was developed that, through a cloud server that will use Natural Language Processing (NLP) and Automatic Translation (AT), transforms the voice into MSL. The signs will be made by a digitally animated avatar on a mobile device .

Keywords: Mexican Sign Language · Application · Natural Language Processing · Database · Inclusive education · Avatar · Hearing impairment

1 Introduction

In Mexico, the number of deaf persons amounts a total of 2.4 million, of which 84,957 are known to be under the age of 14. According to the national survey of demographic dynamics 67% attend school and data of the Sectoral Coordination of Primary Education show that 38,418 attend the 514 primary special education centers.

The sign language is the tool deaf people use to communicate with the world. Mexican Sign Language (MSL) is a natural language with its own vocabulary, semantic and grammatical structures that distinguish it from the rest of the languages, whether spoken or not.

The development of technological tools focused on MSL that facilitate the integration of members of the deaf community it is a task to which we have led our efforts.

Automatic translation from any spoken language to a sign language has several challenges and requires expertise in fields like natural language processing, computer vision and the design of computer graphics.

Natural language processing techniques allow to carry out a linguistic analysis of sign language, identifying its structure and characteristics for creating models. In our case the developed models LET the creation of rules for automatic translation or generation of Mexican Sign Language.

Virtual environment design engines today are capable of creating animations of the human figure articulated enough to represent sign languages. An ideal character should be able to make the necessary movements of the hands, arms, face, head and even the eyes. However, the design of the animations is not enough yet, you must know a set of rules and patterns with which the system can generate a text representation in sign language.

In the case of MSL, it is important to translate using the proper grammar of the language and the appropriate vocabulary, although this not fully regulated due to the lack of a written representation of signs. Other challenge is the display of the sign which includes a lot of movements and body parts.

This work proposes the development of an application that, through a cloud server that uses Natural Language Processing (NLP) and Automatic Translation (AT), transforms the voice into Mexican Sign Language (MSL). The signs will be made by a digitally animated avatar on a mobile device; so that the student can understand more clearly the class taught by the teacher.

2   Automatic Translation into Sign Languages

This work combines sign language translation with avatar technology to improve inclusive education. In this sense, this section explores some works of automatic translation from spoken or written languages to signs. There are different works that approach and propose a different solution to help people with hearing disabilities using machine translation and natural language processing, developed both in national and international universities, as well as associations that they are dedicated to this and have products in distribution in the market.

Automatic translation methods based on corpus is useful for spoken languages. In sing languages most of the works focus on rule-based methods. It has been developed and applied for many languages such as English [15], Arabic [10], French [7], and Spanish [12]. There are attempts to generate rules automatically for French to French Sign Language (FSL) [7], and some other works that explore the statistic translation [2]

It is well known that the Spanish has varieties according to the country and region. San-Segundo, et al. [14] translate from speaking Spanish to Spain Sing Language (LSE: Lengua de Signos Esañola) represented by an 3D Avatar for people applying for Identity Card. In [12] is presented a translation from written Spanish to Mexican Sing Language (LSM: Lengua de Señas Mexicana) represented by sequences of video for restricted grammar structures. For LSM, in [3] is presented a classification of signs with artificial data.

Once the automatic translation process is done, the output is usually a word sequence that should represent by signs. This representation can be done by images, videos, animations or avatars. Because of the advance technology in 3D animation, recent works use avatars to display the signs. Avatars are used in [5] to display translation from written Greek to GSL, from German to Swiss German Sign Language (DSGS) [6], from Spanish to LSE [14], from English to ASL [15] using an avatar with Inverse Kinematic, among others. Some other works, not only use avatars to display the signs: KAZOO [1] allows automatic sign production with a 3D avatar, [11] discusses avatar optimizations that can lower the rendering overhead in real-time displays focused on ASL, and some evaluations in facial expressions of avatars has be done [16].

Applications and algorithms for automatic translations to signs are mainly focus on a topic because of the vocabulary and grammar structures. Using rule-based methods, [14] develop an application for official explanations for identity card for Spain Sing Language (SSL); for bus information, [9] translates to SSL with a 3D avatar animation module. [15] focus on railway station announcements translating to American Sign Language. Most of the topics focus on mobility or public services, missing the educational field.

The Salamanca Declaration (UNESCO, 1994) [4], a political document that defends the principles of inclusive education, recommends that all students have the right to develop according to their potential and develop skills to participate in society. Several international documents, the Salamanca Declaration (UNESCO, 1994) mentioned above, and the Standard Rules for the Equalization of Opportunities for Persons with Disabilities, reflect the need to use sign language as a vehicular language in the education of deaf students.

Sign language is a tool that leads us to interact, communicate, think and learn as well as being part of everyone from a very young age. Therefore, it is necessary to include this in educational programs in order to awaken in society the importance of inclusion for the benefit of all.

3   Translation of Voice to Signs in History Class

A system was developed to help with the communication of people with hearing disabilities in an inclusive environment. In this case for children between 9 and 11 years old who are studying the 4th grade History course at the elementary level. This work was developed for Android smart mobile devices. Figure 1 shows the main processes carried out by the system in order to get successful translations.

3.1   Automatic Translation to MSL

The subject of History has grammatical structures that it implies are simple and in common use, that is: Subject + Verb + Complement, for example:
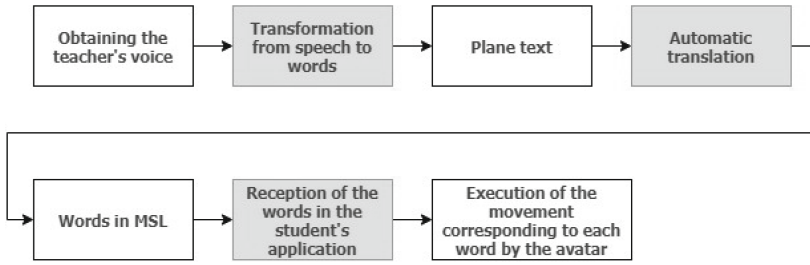
Fig. 1. System block diagram.

"*Miguel Hidalgo started independence of Mexico on September 16, 1810*"
Subject: *Miguel Hidalgo*
Verb: *started*
Completion Begin: *the independence of Mexico on September 16, 1810.*

History uses common vocabulary, with particular names but not as technical as in other subjects, such as mathematics. It is useful because sign language is not very developed in this type of vocabulary, the SEP's dictionary for the deaf mute reaches only 535 signs [8], this means that there are many words to document that are missing, taking into account that the Spanish Royal Academy has approximately 93,000 [13] words.
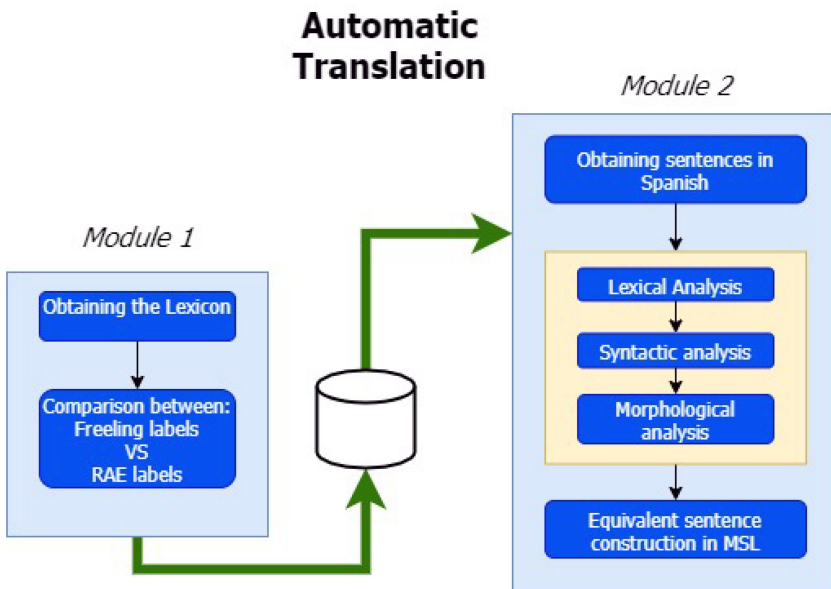

Fig. 2. Interpretation process to restricted vocabulary [12].

Figure 2 shows the representation of the automatic translation. There are two main modules. Module 1 was used to construct the corpus. There were analyzed several common phrases in History classes to get the lexicon. In order to save the vocabulary, there was also done a comparison between grammar labels of the words. Module 2 focus on the translation, 13 sentence structures were studied that resulted in translation rules for 6 syntactic trees, these trees allow the translation of up to 52 grammatical structures. First we decode the speech to sentences in Spanish, then we do a natural language processing by a lexical, syntactic and morphological analysis. We use a rule based method to translate the sentences to a equivalent sentence construction in MSL.

Here is an example of the translation.
The person says: *Children play with the ball.*
It would be interpreted as: *Child many ball play.*
Avatar. *Represent each word through movements.*

3.2   Signal Representation with Avatar

Once the final translation of the sentence entered into the system has been obtained, it will be sent to the student's mobile device, which represents the movements of the avatar thanks to the animation engine Unity Engine. The system will be in constant communication with a server using a Restful API programmed in Flask, which is a Python Framework dedicated to creating Web applications, which is connected to a database that stores all the corpus information, from restricted vocabulary, machine translation rules, as well as synonyms and collocations for this vocabulary in MSL, this data is found in relational tables stored on an SQL server. It should be mentioned that being a prototype and having a restricted vocabulary, the teacher must speak loudly, clearly and at a low speed, omitting words that are outside the restricted vocabulary.

3.3   System Interaction for Automatic Translation

Figure 3 shows the parts that make up the system. The teacher has a mobile device to speak as in a regular History class. He uses the microphone that is integrated in smartphones and tablets, or by means of an external microphone the system will take the teacher's voice to transform it into plain text using Google's speech recognition API.

Once the plain text has been obtained, it is sent by means of HTTP requests to the server, which is interpreted into a restricted vocabulary, performing a lexical, syntactic and morphological analysis (see Illustration 3), a process currently used by the "Direct Translation System of Spanish to MSL with marked rules" [12]; once the restricted vocabulary is obtained, it is automatically translated into MSL, once the words in MSL are obtained, the database is searched to verify the existence of collocations or synonyms, which are replaced by their respective equivalents in MSL.

The avatar representation is show in an mobile device. The interface is focus on students. They can access the class with a key given by the teacher.
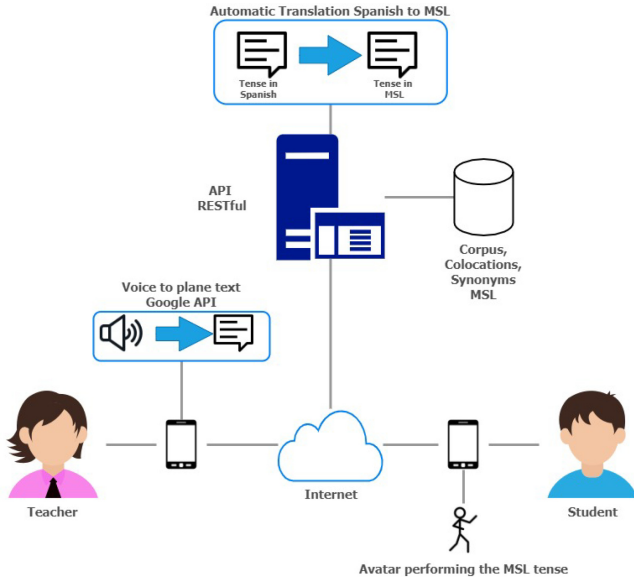
Fig. 3. General diagram of the system.

4   Tests and Results

Figure 4 shows the 3D avatar that the student displays once they enter their email associated with their account and a valid class code. It is here where the student appreciates all the corresponding movements with the AT from Spanish to MSL of each phrase said by the teacher, it should be remembered that the application has 2 parts, teacher and student, as shown in the General diagram of the system.

Fig. 4. 3D avatar.

Table 1. BLEU analysis.

Sentence	Server Response	Time ms	Expected response	BLEU
Â¡Hola, buenos días!	[”hola”, ”buenos-días”]	351	hola buenos-días	1
Â¿Cómo están, niños y niñas?	[”como-estar”,”niño”, ”mucho”,”niño”, ”mujer”, ”mucho”]	438	como-estar niño mucho niño mujer mucho	1
Vamos a iniciar nuestra clase de Historia	[”nuestro”, ”clase”, ”historia”, ”empezar”]	272	nuestro clase historia empezar	0,4347
El tema es El inicio de la guerra de independencia y la participación de Hidalgo y allende	[”tema”, ”empezar”, ”guerra”, ”independencia”, ”participación”, ”hidalgo”, ”Allende”]	617	tema empezar guerra independencia participacion hidalgo Allente	1
Observaremos una imagen de su libro de texto	[”dibujo”, ”libro”, ”ver”]	319	dibujo libro ver	1
Página 166	[”página”, ”num-cien”, ”num-sesenta”, ”num-seis”]	169	pagina num-cien num-sesenta num-seis	1
Página 166	[”página”, ”num-cien”, ”num-sesenta”, ”num-seis”]	169	pagina num-cien num-sesenta num-seis	1
Después contesten las siguientes preguntas	[”después”, ”siguiente”, ”preguntar”, ”mucho”, ”contestar”]	297	después siguiente pregunta mucho contestar	1
Â¿Qué tíulo pondrís a esta pintura?	[”qué”, ”título”, ”dibujo”, ”poner”]	396	qué título dibujo poner	1

Table 1. (*continued*)

Sentence	Server Response	Time ms	Expected Response	BLEU
Â¿Qué están haciendo estos hombres?	["qué", "hombre", "mucho", "hacer"]	412	qué cara explicarr	1
Â¿Qué expresan sus rostros?	["qué", "cara", "explicar"]	321	qué cara explicar	1
Â¿Cómo están vestidos?	["como-estar", "vestir"]	286	como vestir	0,5
Explica porque lo elegiste.	["porque", "explicar", "elegir"]	228	explicar porque elegir	1
Ya vimos en clases anteriores	["ahora", "en", "clase", "mucho", "pasado", "ver"]	309	ahora clase mucho pasado ver	0,53
Recordaremos las causas de la guerra de independencia	["causa", "mucho", "guerra", "independencia", "recordar"]	371	causa mucho guerra independencia recordar	0,53
Anoten lo que recuerden en el pizarrón	["qué", "en", "pizarrón", "anotar", "recordar"]	346	en pizarrón anotar	0,66
Anoten lo que recuerden en el pizarrón	["qué", "en", "pizarrón", "anotar", "recordar"]	346	en pizarrón anotar	0,66
Leeremos la página 168, 169 y 170 del libro de historia	["página", "num-cien", "num-sesenta", "num-ocho", "num-cien", "num-sesenta", "num-nueve", "num-cien", "num-setenta", "libro", "historia", "leer"]	691	página num-cien num-sesenta num-ocho num-cien num-sesenta num-nueve num-cien num-setenta libro historia ver	1
			Average BLEU : 0.87	

While Table 1 shows the comparison of the expected translation against the translation thrown by the system, a python tool called "sentence_bleu" from the Nltk library was used to obtain the BLEU score, Bilingual Evaluation Unserstudy, which is a method of evaluating the quality of translations performed by machine translation systems. A translation has higher quality the more similar it is with respect to another reference, which is supposed to be correct. BLEU can be calculated using more than one reference translation. This allows for greater robustness to size compared to free human translations.

As the system only works with restricted vocabulary to successfully translate a history class, the BLEU assessment was carried out in a total of 20 sentences with a total of 53 words and 27 vowels, it must be taken into account that the system spells any word outside the restricted vocabulary, so 27 vowels were included to be able to interpret words unknown to the system.

4.1   Response Time and Words in the Avatar Vocabulary

Tables 2, 3, 4, 5 and 6 below show the execution times of the main processes carried out by the system, such as AT carried out on the server and animation of the words carried out on the mobile device, as well as the difference in percentage of the "dictation" that would be the equivalent of speaking normally, against the process carried out by the translation system.

Table 2. Response time 1.

Word	Time in seconds					
	AT	Animation	Spell	Total	Total traduction	% difference
Hola	0.59	1.5	1.5	3.59	2.09	39.33%
Cómo están	0.59	1.8	1.5	3.89	2.09	39.33%
Clase	0.59	1.3	1.5	3.39	2.09	39.33%
Historia	0.59	2.88	1.5	4.97	2.09	39.33%
Allende	0.59	2.44	1.5	4.53	2.09	39.33%
Hoy	0.59	3.4	1.5	5.49	2.09	39.33%
Ver	0.59	1.34	1.5	3.43	2.09	39.33%
Tema	0.59	1.34	1.5	3.43	2.09	39.33%
Hidalgo	0.59	2.39	1.5	4.48	2.09	98.67%
Average	0.59	2.043333	1.5	4.1333333	2.63333	75.56%

Table 2 shows a comparison of time between sentences and words that are represented in MSL with a single sign. As you can see in the last column, the signs that are made up of more than 1 movement (History, Allende), are the ones that take the longest time when animating, but in the Automatic Translation process they all take the same time. 0.59 s.

Table 3 shows a comparison of the signs that represent the alphabet, mainly useful for spelling words that are outside the restricted vocabulary. These translations take the same time to process simple phrases and words as the ones exemplified in Table 2, but the animation time is significantly reduced to 0.4 s, but as you can see the time increase is 98%, due to that the translation time added to the animation time far exceed the normal dictation that a personal one would do when speaking.

Table 3. Response time 2.

Word	Time in seconds					
	AT	Animation	Spell	Total	Total traduction	% difference
A	0.59	0.4	0.5	1.49	0.99	98.00%
B	0.59	0.4	0.5	1.49	0.99	98.00%
C	0.59	0.4	0.5	1.49	0.99	98.00%
D	0.59	0.4	0.5	1.49	0.99	98.00%
E	0.59	0.4	0.5	1.49	0.99	98.00%
F	0.59	0.4	0.5	1.49	0.99	98.00%
G	0.59	0.4	0.5	1.49	0.99	98.00%
H	0.59	0.4	0.5	1.49	0.99	98.00%
I	0.59	0.4	0.5	1.49	0.99	98.00%
Average	0.59	0.4	0.5	1.49	0.99	98.00%

As a complement to Table 3, Table 4 shows a set of vowels and consonants, which represent the translation time for words that the system will have to spell when outside the restricted vocabulary. The difference time is less than the spoken time because the system takes very little time to perform the Automatic Translation and the signs of the letters are mostly fast movements and with little body movement. As a result the increase in the time of translation and animation compared to speaking, is only 52.35%.

Table 4. Response time 3.

Word	Time in seconds					
	AT	Animation	Spell	Total	Total traduction	% difference
ABC	0.69	1	1.5	3.19	1.69	12.67%
ABCDE	0.75	2.39	2	5.14	3.14	57.00%
ABCDEFG	0.77	3.19	2.5	6.46	3.96	58.40%
ABCDEFGHI	0.84	4.6	3	8.44	5.44	81.33%
Average	0.7625	2.795	2.25	5.8075	3.5575	52.35%

Table 6 used longer sentences as a reference, in which translations with longer sign lengths were expected, as it can be seen, the time of the Automatic Translation increased considerably on the server as they were more complex grammatical structures as well as the time of execution of the signs on the mobile device so the average percentage difference against a common class was 296.24%

Table 5. Response time 4.

Sentences	Time in seconds					
	AT	Animation	Spell	Total	Total traduction	% difference
Clase de historia de Hidalgo y Allende	6.5	8	6	20.5	14.5	141.67%
Hoy hay clase de historia de Hidalgo y Allende	8	11	6	25	19	216.67%
El tema de hoy es el inicio de la guerra de independencia y la participación de Hidalgo y allende	18	20	7	45	38	442.86%
Hola ¿Cómo estás hoy?	3.4	9.2	3	15.6	12.6	320.00%
El tema de hoy es el inicio de la guerra de independencia	10	13	5	28	23	360.00%
Average	9.18	11.64	5.4	26.82	21.42	296.24%

And as a last test to compare the translation times of the system, long sentences with words outside the system were included in the system restricted vocabulary, as we can see this mainly increased animation time by having to run the words in spelling mode, which increased the total time against a common class by 430.08%

Table 6. Response time 5.

Sentences	Time in seconds						
	Total words	AT	Animation	Spell	Total	Total traduc- tion	% differ- ence
abcde Hay clase de historia abcedefg	2	6.5	8	6	20.5	14.5	141.67%
abcd tema de hoy es la independe- cia	1	4	8.4	3	16.4	12.4	413.33%
abcdefghijk	3	8	18.9	6	35.9	26.9	448.33%
Average	2	6.166	11.766	5	24.266	17.933	430.08%

4.2   Field Tests

Figure 5 shows the field tests carried out in "La casa del Sordo", there you can see an instructor of the "Casa del Sordo" evaluating the precision of signs and the translation that was done by the system. In the image he is doing the sing of Ignacio Allende, a leader in Mexican Independence.

Fig. 5. Field tests 1.

5   Conclusions and Future Work

Sign language presents various peculiarities, since it is adaptable to different situations, moments and places; therefore, the study is complicated, as is the creation of a mobile application to translate the Mexican sign language. Because to carry out a translation to MSL of the central zone it is not possible to generate word by word, since the sentences can be represented in multiple ways depending on the region, zone, country, it was concluded that in this case they will be represented with that of Mexico City. An algorithm capable of simulating a grammatical tree can be created, which is in charge of generating word arrangements to translate, however, it must be executed faster to achieve a more realistic translation. The processing speed of the text coming from the voice, together with the reproduction of each animation, exceeds for a considerable time the human voice and translation. Natural language processing techniques make it easy to categorize the words that make up a sentence. The open source Freeling language natural processing library facilitates the processing of all the sentences put in this work, in this way, it was possible to establish a limited vocabulary to teach a history class of 4th grade of primary school. The results obtained after the development and implementation of this work, the scientific advance to help the inclusion of deaf people in education, the incorporation of increasingly sophisticated and accurate software, the growing expectation that users place on technology to supporting the solution of problems in your life, allows us to suggest, as a future work, the incorporation of the following functionalities into the System like the incorporation of more sign animations, that is, to strengthen not only history classes, but all kinds of conversations and situations, improve the processing time of machine translation, and design signs with greater precision and better animation technologies.

References

1. Braffort, A., et al.: KAZOO: a sign language generation platform based on production rules. Univ. Access Inf. Soc. **15**(4), 541–550 (2016)
2. Bungeroth, J., Ney, H.: Statistical sign language translation. In: Workshop on Representation and Processing of Sign Languages, LREC, vol. 4 (2004)
3. Cervantes, J., et al.: Clasificación del lenguaje de señas mexicano con SVM generando datos artificiales. Revista vínculos **10**(1), 328–341 (2013)
4. Domínguez, A.B.: Servicio de información sobre discapacidad. https://sid.usal.es/idocs/F8/ART11921/educacion_para_la_inclusion_de_alum_sordos.pdf. Accessed 2019
5. Efthimiou, E., et al.: From grammar-based MT to post-processed SL representations. Univ. Access Inf. Soc. **15**(4), 499–511 (2016)
6. Ebling, S., Glauert, J.: Building a Swiss German sign language avatar with JASigning and evaluating it among the deaf community. Univ. Access Inf. Soc. **15**(4), 577–587 (2016)
7. Filhol, M., Hadjadj, M.N., Testu, B.: A rule triggering system for automatic text-to-sign translation. Univ. Access Inf. Soc. **15**(4), 487–498 (2016)

8. Hernández, M.T.C.: Diccionario Español-Lengua de señas Mexicana (DIELSEME), SEP, México
9. López-Ludeña, V., et al.: Translating bus information into sign language for deaf people. Eng. Appl. Artif. Intell. **32**, 258–269 (2014)
10. Luqman, H., Mahmoud, S.A.: Automatic translation of Arabic text-to-Arabic sign language. Univ. Access Inf. Soc. **18**(4), 939–951 (2019)
11. McDonald, J., et al.: An automated technique for real-time production of lifelike animations of American Sign Language. Univ. Access Inf. Soc. **15**(4), 551–566 (2016)
12. Pichardo Lagunas, O., Partida Terrón, L., Martínez Seis, B.C., Alvelar Gallegos, A., Serrano Olea, R.: Sistema de traduccion directa del español a LSM con reglas marcadas, Ciudad de México: IPN (2016)
13. RAE, Real Academia Española, Real Academia Española. http://www.rae.es/. Accessed 2019
14. San-Segundo, R., et al.: Speech to sign language translation system for Spanish. Speech Communi. **50**(11-12), 1009–1020 (2008)
15. Shaikh, F., et al.: Sign language translation system for railway station announcements. In: 2019 IEEE Bombay Section Signature Conference (IBSSC). IEEE (2019)
16. Smith, R.G., Nolan, B.: Emotional facial expressions in synthesised sign language avatars: a manual evaluation. Univ. Access Inf. Soc. **15**(4), 567–576 (2016)