Automatic Translation of Sentences to Mexican Sign Language: Rule-based Machine Translation and Animation Synthesis in Avatar

Bella Martinez-Seis[1], Obdulia Pichardo-Lagunas[1,*], Eliot Hernández-Morales[1],
Oscar Rivera-Rodríguez[1], Sabino Miranda[2]

[1] Instituto Politécnico Nacional,
Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas,
Mexico

[2] Universidad Autónoma de la Ciudad de México,
Mexico

{bcmartinez, opichardola}@ipn.mx, sabino.miranda@uacm.edu.mx

Abstract. Sign Languages are mainly used by deaf people. The translation between Spanish of Mexico and Mexican Sign Language is a current challenge that remains unresolved. This paper considers two main areas for a proper translation: automatic translation and sign representation. The first one considers the syntactic of the language. The second one includes the representation of sequential signs. We propose a tool to translate sentences from written Spanish to Mexican Sign Language considering the syntactic from both languages. We use automatic translation based on rules because of the lack of a big corpus. The BLUE score for the translation was about 0.8061, which suggests a good translation. To display the signs, we used a 3D humanoid avatar. Signs Languages are agraphia, so we use a configuration matrix to describe them. We propose a process for Sign Language Synthesis. It takes the configuration matrix of each sign and generates animation rules describing the whole movement and positions that the avatar follows to produce the signs. It allows to increase the signs that the avatar represents easily.

Keywords. Sign languages, automatic translation, avatar, animation synthesis.

1 Introduction

Sign languages are based on manual expressions and various facial, arm, and body movements.

In 2005, Mexican Sign Language (MSL) obtained legal recognition in Mexico, being considered part of the linguistic heritage of the Mexican nation [1]. The MSL has its own syntax, grammar, and lexicon. One or more signs can represent each word or expression in Spanish. It is not a word-by-sign translation and these signs are not the spelling of the words.

Multimedia tools could be of great support for the representation of signs, however, there is an ideological background, research, and knowledge of the grammar of sign language and its custodians, the deaf community, that permeates its development and that it is evidenced by the treatment given to both languages (oral and sign) [2].

The sign is made up of various manual and non-manual elements, they are classified [3] into iconic kinesis, deictic, intermediate, and arbitrary signs. This work focuses on iconic kinesis signs, which reconstruct the object they represent with their hands, body, and space through its shape, movement, or spatial relationship.

These signs are represented in dictionaries like the first Dictionary of Mexican Sign Languages of Mexico City [4] with more than 1000 signs. Sign languages are agraphia, so this dictionary considers the phonology of the MSL of [11] which proposed a segmental matrix, an articulatory
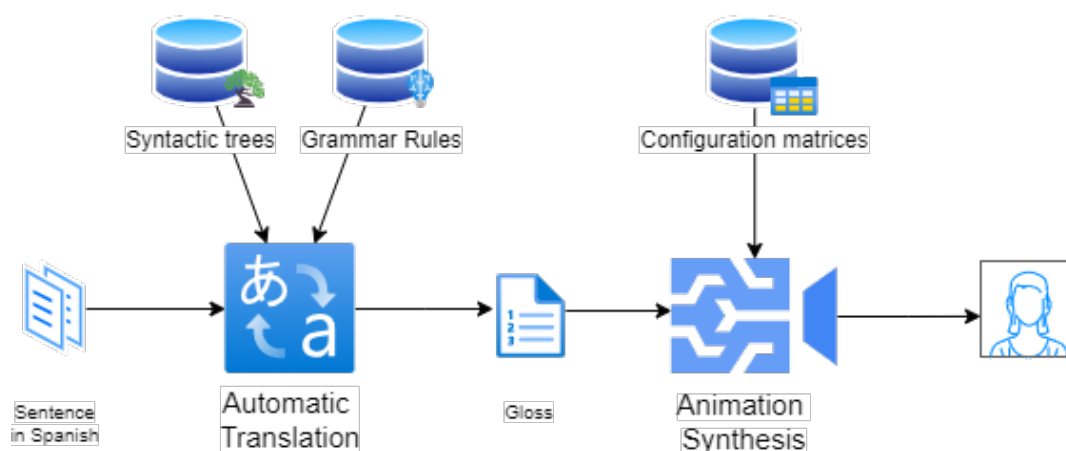
Fig. 1. Arquitecture for automatic translation to MSL with animation synthesis in avatar
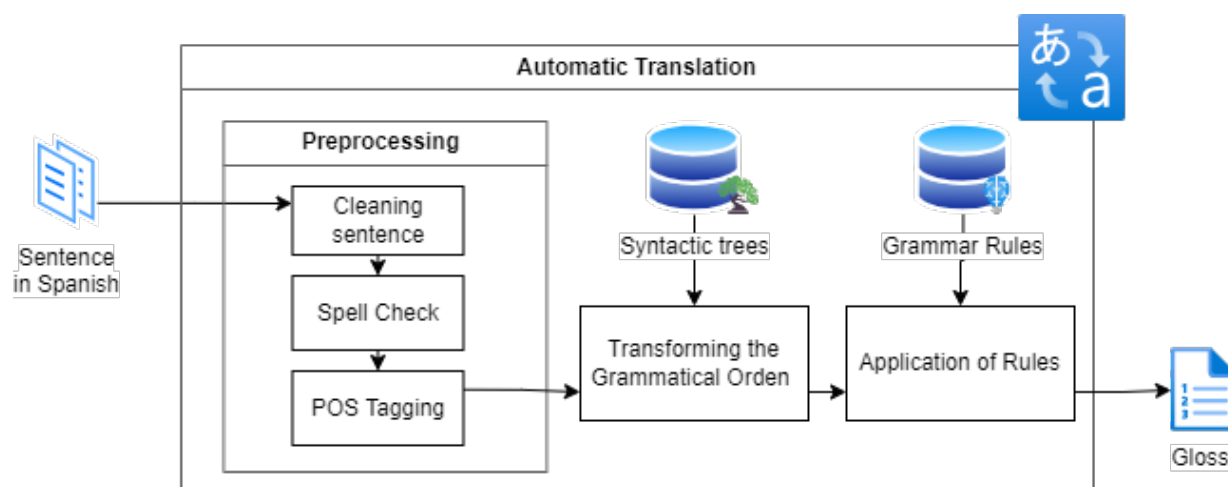

Fig. 2. Translation process.

matrix, and non-manual traits matrix; those matrices will be called as configuration matrix. The representation of signs has been carried out through images, image sequences, videos, and avatars. Recently, there has been an inclination towards systems that generate photo-realistic signs considering the pose of the skeleton.In those works the animation is configured for each of the signs. The biggest drawback is that they must animate sign by sign, which limits the system's potential output growth.

Another problem is the transition between signs. Most of the systems translate words, but if we translate a sentence we will have a sequence of signs, where the transition from one to another must be fluid and not jump between the end of one sign and the beginning of the other.

In this sense, [8] combines automatic translation, body gesture animation, and facial avatar generation for Chinese Sign Language using more than 2000 gloss motions from professional sign language practitioners with motion capture devices.

The computational cost is high even when they use a neural network based on Transformer for the transitions. The tasks involved in the Machine Translation of Spanish of Mexico to LSM are addressed in two main proposals of

Table 1. MSL sentence structure

Spanish	Los	niños	juegan	pelota
	Noun Phrase		Verb Phrase	
	Article	Noun	Verb	Noun
MSL	niño	muchos	pelota	jugar
	Noun Phrase		Verb Phrase	
	Noun	Adverv	Noun	Verb
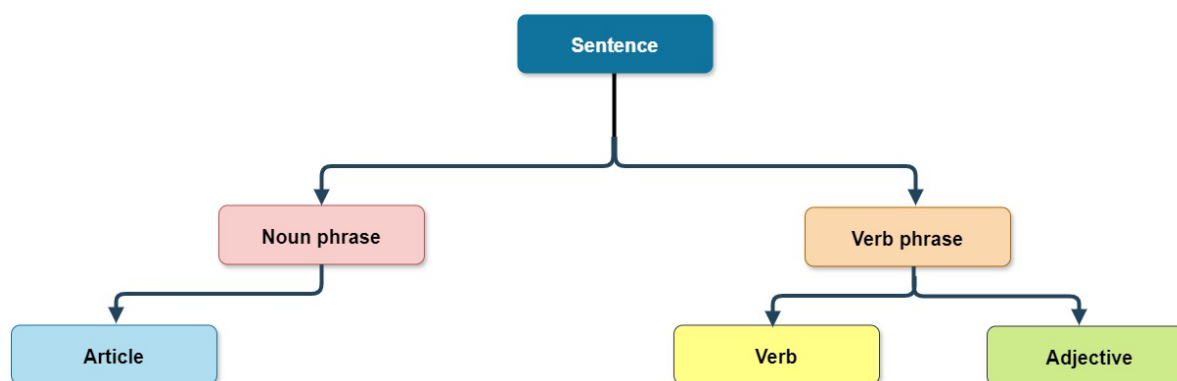
Fig. 3. Intermediate tree (Article, noun, verb and adjective)

this work: automatic translation considering the syntax and the representation of the sign by an avatar with synthesis of animations without previously created animations. This work is, also, an effort to minimize the gap between Mexican Spanish and Mexican Sign Language from a technological perspective.

1.1 Automatic Translation

Machine translation uses machine learning and natural language processing techniques to analyze and convert text or speech from the source language to the target language. without direct human intervention. Traditional machine translation systems were built based on rules, however, modern machine translation uses deep learning algorithms and large amounts of data.

This project uses rule-based techniques for translation since the implementation of modern techniques requires a large amount of data that the MSL does not have.

1.2 Animation Generation for Sign Language

The character animation synthesis method is a virtual character motion in a virtual environment, the most used is physics-based character animation synthesis [6], resulting in the trajectory planning and classical control methods. In recent years, several schemes for sign language production have been proposed, mainly using avatars.

Most of the works that used avatars, generate the animation of the signs, it is a difficult task that requires a complex and steep learning curve. To mitigate it [7] presents a tool to animate an avatar for Sign Languages where the user only defines keyframes and the system interpolates between them; it was tested in six sentences.

But these generated data cannot be expanded, and professional knowledge is required [9, 8]. Generative Adversarial Networks (GAN) combined with graphical techniques has also been used [10] for German Sign Language, it required a big
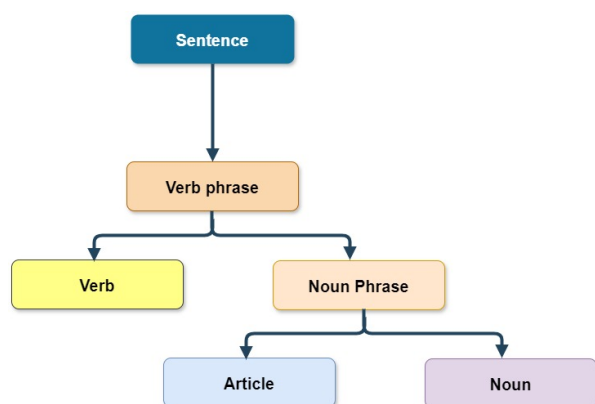
Fig. 4. Simple tree (Article, verb and adjective)

data set that was essential for the robustness and flexibility of their method.

2 Proposed Architecture

The translation between Spanish of Mexico and Mexican Sign Language proposed has been approached from two main areas: automatic translation and sign representation. Figure 1 shows the general proposed solution. The input is a sentence in Spanish.

We used automatic translation based on rules using syntactic trees to translate the sentence into the gloss, or words that represent the signs (see Section 2.1). This gloss will be signed by the avatar, to do so, an animation synthesis process is proposed considering the configuration matrices of each sign (See Section 2.2).

2.1 Module of Automatic Translation based on Rules for MSL

Rule-based machine translation (RBMT) is a method that employs linguistic rules to analyze and translate text. This approach operates on the premise that linguistic rules can be utilized to understand the structure of a language and produce accurate translations.

In RBMT, linguists and translation experts develop a comprehensive set of rules that delineate the grammar and structures of both the source and target languages. These rules are applied to analyze the source text and generate the

corresponding translation in the target language, to do so, we used syntactic tree. Finally, we apply some rules such as eliminating articles or grammatical tenses (see Figure 2).

2.1.1 Preprocessing

The automatic translation module includes a Preprocessing stage of the text written by the user. It is important to know that sentences cannot contain more than 50 characters. This tasks include:

– Correcting spelling errors by removing special characters and punctuation marks whether entered by mistake or as part of the sentence.

– The words written in the interface are compared with a dictionary from the Royal Spanish Academy (RAE) that has 600 thousand words that is housed in a text file.

  For the word search, a database was implemented with SQLite from the text file and the Levenshtein distance comparison algorithm was implemented for search optimization.

  The word with the smallest distance is the suggested correction. Finally, when all the words have been verified, the original words are replaced with the suggested ones.

– Once the sentence is preprocessed, SPACY is used to POS tagging of the sentence to identify the characteristics of its components. This result is stored in JSON format.

2.1.2 Translation Considereing Grammatical Order

The implementation of the translation will be based on the following principles: There are two formal tools for communication with the signing community in Mexico: Mexican Sign Language (MSL) and Signed Spanish (SE).

The MSL, like any language, has its own syntactic and semantic structures, as well as its own lexical components. Unlike the MSL, signed Spanish is based on the components of the dominant language, in this case Spanish, and only interprets it word by word. Spoken languages
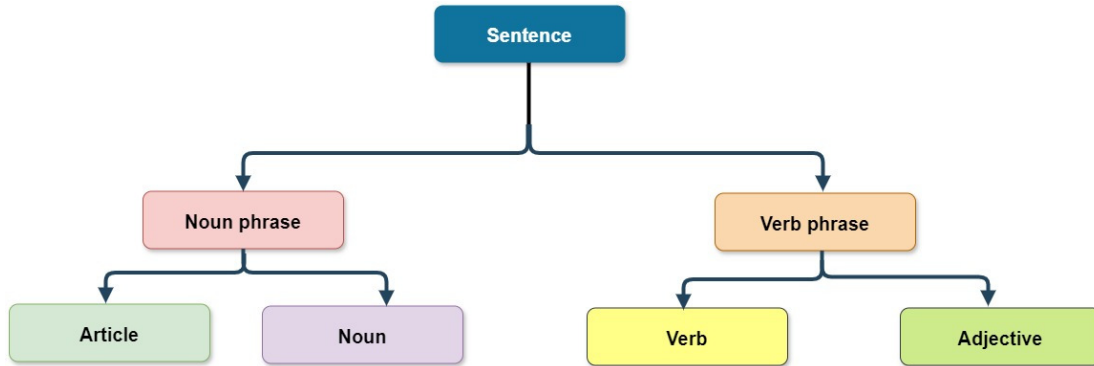
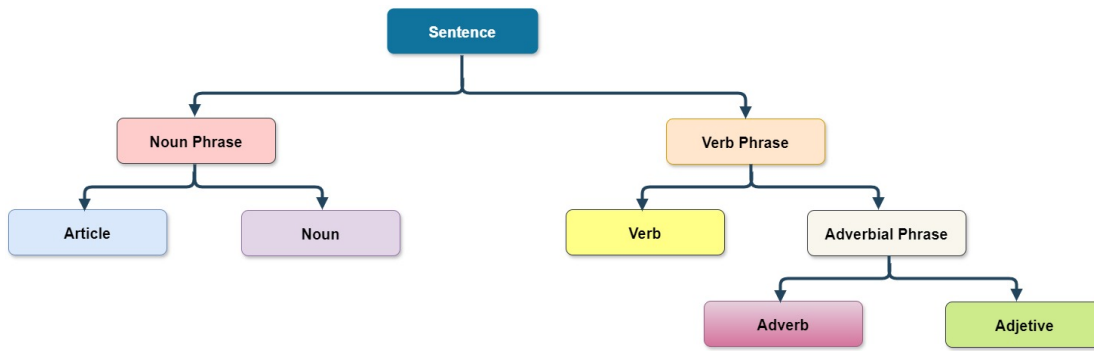**Fig. 5.** Complex tree (Two articles, noun, verb, article and subject)

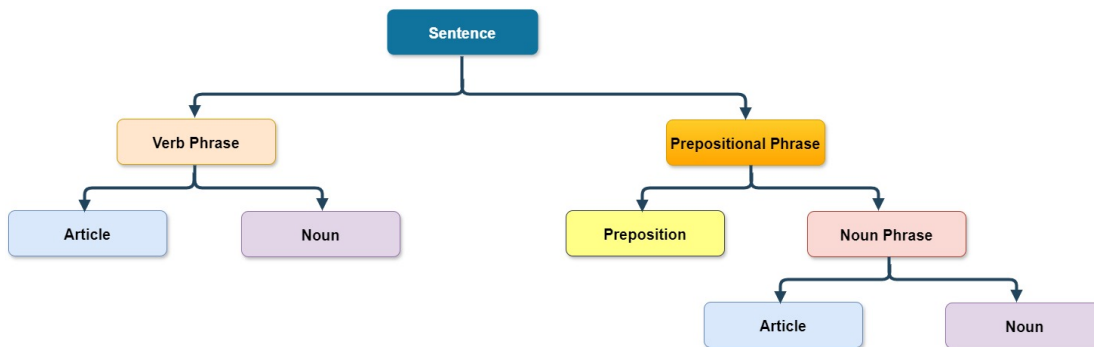**Fig. 6.** Simple tree (Verb, article and noun)

Fig. 7. Simple tree (Verb, article and noun)

have a sequential order, while, in MSL, movements can be performed simultaneously. The order of the elements that make up a sentence can be different, for example.

The elements that make up a sign are not only those represented by the hands, they can also include other parts of the body, facial expressions, and position of the body in space.

2.1.3 Syntactic Trees

The MSL also has a set of valid syntactic structures to form a sentence, this allows the construction
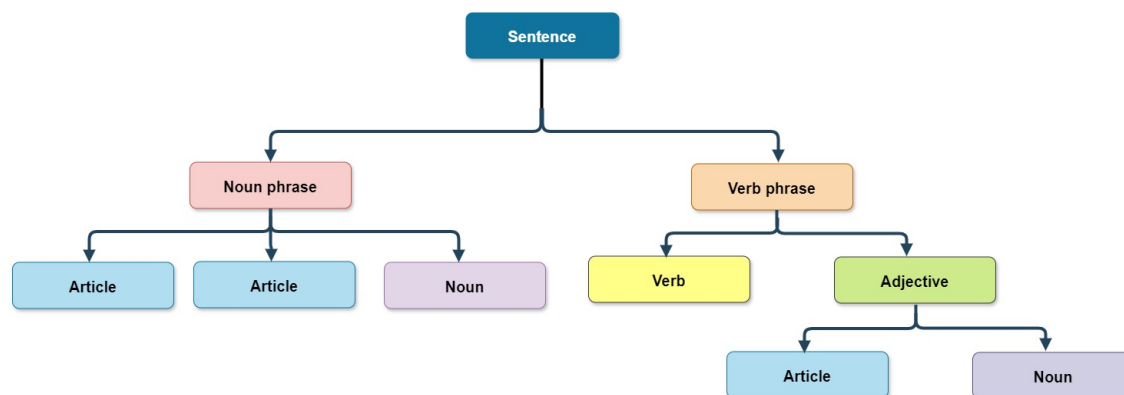
Fig. 8. Intermediate tree (Article, noun, verb, adverb and adjective)

of different trees, which can be used to validate any sentence entered into the system. Each syntactic tree was classified according to the difficulty of its reading.

Therefore, trees with less than 3 components are considered simple; those with between 4 and 5 components are considered intermediate, and those with more than 5 components are considered complex. Our system used six basic syntax: simplex tree: Figures 3, 4, intermediate tree: Figures 5, 7, 6, and complex tree: Figures 8.

2.1.4 Transforming the Grammatical Order

Once we have the tagged sentence, the program reads and compares the sentence with the syntactic trees. If the syntactic is considered in the syntactic trees of the previous section, the translation will be in Mexican Sign Language (MSL), otherwise, it will be in Signed Spanish (SE). For the translation to MSL, the syntactic tree is read in a different order where most of the time the verb goes at the end.

2.1.5 Application of the Grammar Rules

MSL has fewer signs than words in Spanish, then some grammatical elements are used like they are, eliminated in the translation, or transformed according to the following rules:

– Just 13 prepositions are used like they are.

– Some personal pronouns and possessive adjectives are available.

– Indefinite and demonstrative adjectives are eliminated.

– 14 adverbs of time are considered.

– 7 adverbs of grade are available and 3 more are transformed, for example "tan" (such) is transformed to the gloss of "equal".

– 17 adverbs of place are used, the others are eliminated.

– 7 adverbs of manner are used.

– Adverbs of affirmation "yes" and "true" are used.

– Adverbs of negation "no" is used and others like "nothing", "neither", "never", and others are transformed to "no".

– Adverbs of doubt are removed.

– The basic possessive pronouns are considered.

– Demonstrative pronouns are not considered.

– Indefinite pronouns are not considered.

– Words that can be used in feminine are transformed to the word plus the sign of "women", like "girl" is transformed to "boy"+"women".

– The sign "many" is added if the word is in the plural.

– Tenses are not considered.

Representation	Feature	Description	Image
✚	Open	Proximal joint and distal joints extended.	
^	Flattened	Proximal joint closed and distal joints extended.	
"	Hook	Proximal joint extended and distal joints partially closed	
▬	Close	Closed proximal and distal joints	
O	Relaxed	Relaxed joints	

Fig. 9. Finger posture traits of the configuration matrix

Matrix	Component		Example: GOOD (bien)	
Segmental Matrix	Movement M or Detention D		D-------------M----------------D	
	Movement		Lin	
Articulatory Matrix	CM Configuration of the fingers		1234+/a+	1234+/a+
	UB Location	Point SM REL LOC	GemD1D2D3 Cont - Lab	GemD1D2D3 Cont Inf Lab
	DI Direction		Palm	Palm
	OR Orientation		Neut	Neut
No-manual Matrix	Face Head Body		∅	∅

Fig. 10. Configuration matrix with the Representations (abbreviations) and an example of the sign GOOD

SEG	Detention														
Word	CM (phalanges)													UB	OR
	1A	1B	1C	2A	2B	2C	3A	3B	3C	4A	4B	4C	a	GemPol-Cont-AbjMe	Neut
Nothing	90°	90°	90°	90°	90°	90°	90°	90°	90°	90°	90°	90°	0°	COD	HOM
														180°	ROT 45° · Z- 45°

Fig. 11. Representation of the angles of the sign nothing (nada)

2.2 Animation Synthesis for Signs from Configuration Matrix

The gloss is the ordered set of words that will be represented in signs. Each word has a written representation given by the configuration matrix. We got the written representation of about 1600 signs, some of them given by our observation. Nevertheless, there are glosses that we did not consider then, the word breaks down into letters and their own configuration matrix.

The movements that the avatar does, are gotten from the configuration matrix through the animation synthesis that will be describe in this section. Traditional animation synthesis focuses on the skeleton, nevertheless, we require details in hands. Four of the fingers of the hands have three phalanges which give us the five possible positions, shown in Figure 9, for each finger.

In addition, there are interaction traits between fingers like a progressive close, separation, or cross. The system is difficult to control due to the precision required by the signs in the finger, wrist, elbow and shoulder joints, and the coupling of the variables of each joint. Which makes the proposed animation synthesis very relevant. Our proposal decomposes the sign into small parts based on the configuration matrix. It includes [11] a segmental matrix, an articulatory matrix, and a no-manual traits matrix.

– The segmental matrix describes if the signs is in detention or movement. The possible movements are described as contour movement (lin-lineal, circ-circular, ...), local movements (rot-wrist rotation, rsc-scratching, ...), temporal quality (rap-fast, sost-sustained,...), no temporarl quality (amp-extended, tns-tense, ...), contact (roz-rubbing and reb-rebound), and espacial (PH-horizontal plane, PS-surface plane, ...).

– In the articulatory matrix describe the configuration of the fingers (manual configuration), the location, direction, and orientation.

– In the manual configuration, each finger is represented with a number. Each position has a representation, for example for the four fingers the representation is shown in Figure 9. For the thumb, the lower phalanx could be aligned (a) with the palm (on the side) and opposite (o) to the palm (on the palm), the other positions are: open (+), flatten (∧), or close (-) plus contact between fingers.

– The location (UB) considers the points (Point) of the hand that could define locations, the hand surface (SM) refers to what part of the active hand is facing or in contact with the location, spatial relationship (REL), and location (LOC) that refers to the passive articulator in a location on the body, such as in the hand or in the pointing space.

– Direction (DI) indicates which location the hand and part of the hand are directed to the body or the plane of the surface.

– The Orientation (OR) refers to the horizontal plane.

– The no-manual traits includes other parts of the body but it was not consider in this proposal.

Fig. 12. Graphic interface of the avatar that interprets the signs given the gloss from translation

Figure 10 has the representation of the configuration matrix which has a lot of abbreviations given by the author of that matrix [11], all of them are not described in this paper but have been considered in the proposal.

For example, for the sign of GOOD (bien in Spanish) the segmental matrix shows that it has two detentions (D), the one that shows the initial position of the sign and the one that describes the final position of the sign that the person in the figure does. Those detentions are connected by a movement M wich is linear.

The articulatory matrix describes for the manual configuration (CM) that the fingers 1,2,3, and 4 are open (+) and the thumb has an aligned (a) and open (+) position.

The location (UB) means that at first the fingertips (Gem) make contact (Cont) with the lips (Lab), then the hand is then directed towards the front and below (Inf) this location. The direction (DI) describes that the palm (Palm) is in front of the signer's body throughout the entire movement.

And a neutral (Neut) orientation to the horizontal plane. In the example, there is just one hand involved but we consider the dominant and no-dominant hand. The configuration matrix of the considered signs is represented in a documental database.

It is important to note that the words entered into the current database have been considering the team's interpretation and understanding based on existing matrices and some other sources. This database allows anybody, with the configuration matrix of the sign, could add a new sign to this database.

Then, each small part of the decomposed sign is associated with joints of the skeleton of the avatar. Each part has an angle (See Figure 11) and then is is represented by a position in the axes X, Y, and Z.

The creation of the avatar involved modeling the body using several figures, the modeling of the face started from a simple base mesh, then extrusion and sculpture techniques were applied.

Subsequently, a refinement is made for fine details such as the fingers. Then smoothing is applied and finally textures are applied. To give movement to the avatar, a skeleton was used, known as rigging. The avatar was developed in Unity where there is a target and hint. The first is involved in the translation or displacement of the joint and the second in the rotation and flexion of the joint. The meticulous distribution of targets and hits allowed for naturalness in the movements.

Movements are achieved by calculating trajectories with linear interpolation, Lerp. Given the complexity of the movements and the finesse required, rotations must be managed with precision and consistency, which is why we focus on the correct alignment and orientation of the joints, controlled by the Euler angles. Therefore, rotations were made between the dice and Euler angles.

3 Evaluation of the Translation

3.1 Test Data Sets

The Manual de gramática de la Lengua de Señas Mexicana has all the rules and examples of how to translate it into LSM. A limitation of this validation process is that the references associated with each sentence did not cover all the structures validated in the system. We used 57 sentences from that book.

3.2 Metric

For the evaluation of the tanslation, we used the metric BLEU. BLEU [12] is a metric that is widely used to evaluate Natural Language Processing (NLP) systems that produce language, especially machine translation (MT). It is a weighted average of variable length phrase matches against the reference translations. The cornerstone of BLEU is the precision measure. And it could be evaluated by n-grams. The equation of BLEU is:

$$\text{BLEU} = \text{PB} \cdot \exp\left(\sum_{n=1}^{N} w_n \log P_n\right), \quad (1)$$

where each n-gram has a weight w_n such that $\sum_{n=1}^{N} w_n = 1$ and the Penalty for brevity PB.

3.2.1 Results

We evaluate BLEU with bigrams. We compare our translation to the ones given in the sentences of the book. The average BLEU score, which was approximately 0.8061. This score is relatively high and suggests a good correspondence between the generated translations and the references used to evaluate them. A score above 0.5 is generally interpreted as good translation quality,

4 Graphic Interface Validation

The validation of the graphic representations of the signs was evaluated by an certified LSM interpreter. Some of the comments received were around gratitude for considering the grammatical structure of their language and not representing only signed Spanish. Among the aspects to improve, he commented that he considers it can be a support tool for learning more than an interpreter due to the speed and lack of signs that consider the spatial environment.

Figure 12 shows the graphic interface where the typed sentence was an equivalent of: "you has to claen the car" (Hay ke limpair el auto) where there are orthographic and typing errors. First, the system preprocesses the sentence and the final sentence to translate eliminates those errors giving "you have to clean the car" (Hay que limpiar el auto). The translation to LSM are the gloss "car clean need" (auto limpiar necesitar). In the moment of the screenshot, the avatar is showing the sign of "need" (necesitar) which is the word that could be read in the top of the figure.

5 Conclusion and Future Work

A system was implemented that allows the visualization of the automatic translation of written Spanish into Mexican Sign Language. The translation process was carried out using a pre-established set of syntactic trees and previously identified semantic norms, obtaining a score of 0.8031 in the BLEU metric, which suggests a substantial correspondence between the translations and the references. While the visualization of the translation was carried out,

it was carried out with the implementation of a graphical interface that, with the help of a dynamic Avatar system and the use of sign configuration matrices, allowed the dynamic reproduction of the signs that make up the translation. prayer.

Acknowledgments

References

1. **Cámara de Diputados del H. Congreso de la Unión (2011).** Ley general para la inclusión de las personas con discapacidad. Secretaría General, Secretaría de Servicios Parlamentarios.

2. **Cruz-Aldrete, M. (2014).** Hacia la construcción de un diccionario de lengua de señas mexicana. Revista de Investigación, Vol. 28, No. 83, pp. 57–80.

3. **Acosta, L., Calvo, T., Maya, D., Sanabria, E. (2004).** Diccionario español-lengua de señas mexicana. DIELSEME, México: Dirección de Educación Especial SEP.

4. **Mercader-Flores, C.A., Dellamary, L.E., Ramírez, M.R., Pool, M., Cruz-Aldrete, M. (2017).** Diccionario de lengua de señas mexicana de la Ciudad de México.

5. **Delorme, M., Filhol, M., Braffort, A. (2009).** Animation generation process for sign language synthesis. Proceedings of the Second International Conferences on Advances in Computer-Human Interactions, pp. 386–390. DOI: 10.1109/achi.2009.29.

6. **Ma, X. (2016).** Study on synthesis method of character animation under the perspective of physics. Proceedings of the International Conference on Economy, Management and Education Technology, pp. 1125–1130. DOI:10.2991/icemet-16.2016.245.

7. **Cabral, P., Goncalves, M., Nicolau, H., Coheur, L., Santos, R. (2020).** PE2LGP Animator: A tool to animate a portuguese sign language avatar. Proceedings of the Language Resources and Evaluation Conference and the 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives, pp. 33–38.

8. **Hu, L., Li, J., Zhang, J., Wang, Q., Zhang, B., Tan, P. (2022).** A speech-driven sign language avatar animation system for hearing impaired applications. Proceedings of the 31st International Joint Conference on Artificial Intelligence, pp. 5912–5915. DOI: 10.24963/ijcai.2022/852.

9. **Ebling, S., Glauert, J. (2013).** Exploiting the full potential of jasigning to build an avatar signing train announcements. Proceedings of the Third International Symposium on Sign Language Translation and Avatar Technology, pp. 1–8. DOI:10.5167/uzh-85716.

10. **Stoll, S., Camgoz, N.C., Hadfield, S., Bowden, R. (2020).** Text2sign: Towards sign language production using neural machine translation and generative adversarial networks. International Journal of Computer Vision, Vol. 128, No. 4, pp. 891–908. DOI: 10.1007/s11263-019-01281-2.

11. **Aldrete, M.C. (2008).** Gramática de la lengua de señas mexicana. PhD Thesis, El colegio de México.

12. **Papineni, K., Roukos, S., Ward, T., Zhu, W. (2001).** Bleu: A method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 311. DOI: 10.3115/1073083.1073135.