LSEAvatar: An Avatar for Spanish Sign Language

María Pilar Agustín-Llach[1], Vanessa Alvear[2,3], César Domínguez[2], Manuel García-Domínguez[2], Jónathan Heras[2], Félix Lanas[2], Gadea Mata[2], Pablo Munarriz-Senosiain[2], Pablo Ochoa[2] and Mirari San Martín[2]

[1]Departamento de Filologías Modernas, Universidad de La Rioja, Spain
[2]Departamento de Matemáticas y Computación, Universidad de La Rioja, Spain
[3]Innovación Riojana de Soluciones IT S.L., Logroño, Spain

Abstract

Approximately 70,000 individuals use Spanish Sign Language (LSE) as their primary means of communication. However, due to the limited prevalence of sign language proficiency among the general population, deaf individuals often face significant challenges in various environments. Therefore, the development of technological systems that facilitate communication between deaf and hearing individuals is essential. In the LSEAvatar project, we address how to translate messages from Spoken Spanish into LSE to facilitate communication for deaf individuals. To achieve this goal, we will employ natural language processing techniques along with deep learning models that convert audio or text into LSE glosses. Ultimately, the project aims for these glosses to be interpreted by an avatar, enhancing access to information and communication for deaf individuals. This project has the collaboration, advice, and validation of the Association of Deaf of La Rioja.

Keywords

Spanish Sign Language, LSE, Glosses, Avatar

1. Introduction

According to the World Federation of the Deaf, approximately 70 million people worldwide are deaf [1]. In Spain, the population with hearing disabilities amounts to around 1,230,000 people [2], with approximately 70,000 individuals [3] using sign language as their primary means of communication. However, due to the limited prevalence of sign language proficiency among the general population, deaf individuals often face significant challenges in various environments, making daily interactions difficult, particularly in the absence of interpreters for translation assistance [4]. Given that sign language is the primary mode of communication for the deaf community, the development of technological systems that facilitate communication between deaf and hearing individuals is essential [5]. Specifically, tools are needed to enable hearing persons to understand signed messages and to help them get their oral messages across in signed modality so that deaf individuals can apprehend them. The present project focuses on the latter aspect.

Currently, one of the most widely used tools allowing deaf individuals to access spoken Spanish, from now on LOE (which stands for *Lengua Oral Española*), is the use of transcriptions or subtitles available in television programs and films. Additionally, applications such as Google's real-time transcription tool [6] can be used to generate text transcriptions that deaf individuals can read. However, there are several reasons why deaf individuals may prefer using Spanish Sign Language (from now on, LSE, which stands for *Lengua de Signos Española*) rather than reading written text. Firstly, many deaf individuals, especially those who are deaf from birth, prefer accessing information through LSE since it is their native language. Secondly, some individuals may struggle to read subtitles at the speed at which they change on the screen. Finally, subtitles generally fail to capture other aspects of oral language such as intonation patterns, volume, rhythm, specific accents and so on.

The aforementioned considerations have led to the establishment of the following objective for the LSEAvatar project. In this project, we aim to develop an avatar capable of translating LOE into LSE using Natual Language Processing (NLP) and Computer Vision techniques. To achieve such a goal, the following specific objectives have been outlined:

- *Datasets*: Collection of diverse datasets to train the models used by the avatar. Specifically, a repository of signed video clips will be built using open-source LSE dictionaries. Additionally, a dataset will be created to convert LOE into glosses — an intermediate written representation of sign concepts.
- *Models*: Development and implementation of various machine learning and deep learning models to support different components of the avatar. These include computer vision models for capturing facial, arm, and hand movements necessary for signing, and rendering these movements onto the avatar; and language models for transcribing audio into text (that is, Speech to Text technology), and converting written LOE into LSE glosses.
- *Integration*: Design and implementation of the avatar, ensuring seamless integration with the models so that the avatar can generate sign language output from spoken or written input.
- *Validation*: Evaluation of the avatar's effectiveness by members of the Association of the Deaf of La Rioja.
- *Deployment*: Implementation of the avatar on various platforms, such as mobile applications and web pages, and exploring its potential integration into videos as an alternative to subtitles.

The development of this avatar will represent a significant step forward in fostering inclusion and integration

SEPLN 2025: 41st International Conference of the Spanish Society for Natural Language Processing, Zaragoza, Spain, 23-26 September 2025.
† Authors listed in alphabetical order
maria-del-pilar.agustin@unirioja.es (M. P. Agustín-Llach); vanessa.alvear@irsoluciones.com (V. Alvear); cesar.dominguez@unirioja.es (C. Domínguez); manuel.garciad@unirioja.es (M. García-Domínguez); jonathan.heras@unirioja.es (J. Heras); felix.lanas@unirioja.es (F. Lanas); gadea.mata@unirioja.es (G. Mata); pablo.munarriz@unirioja.es (P. Munarriz-Senosiain); pablo.ochoa@unirioja.es (P. Ochoa); miren.san-martin@unirioja.es (M. San Martín)

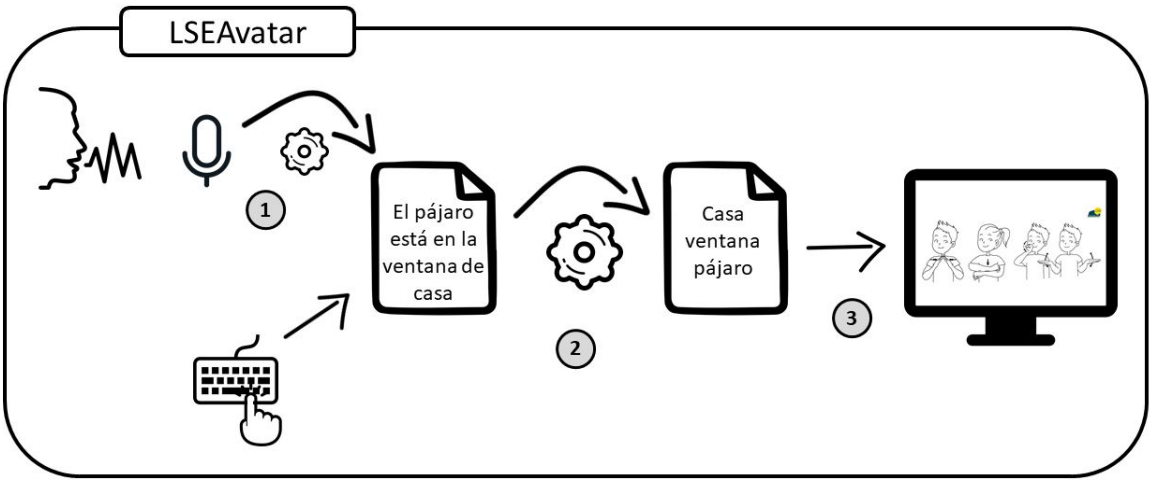


Figure 1: Architecture of the LSEAvatar project. (1) LOE audio to LOE text (in the example of the figure the text says "The bird is on the window of the house"); (2) LOE text to LSE glosses (in the example of the figure the glosses are "House window bird"); and (3) LSE glosses to avatar.

for deaf individuals. At present, the only existing avatar with a similar purpose is the one presented in [7, 8], which focuses solely on signing individual words using the LSE alphabet. As such, it does not incorporate specific signs or the grammatical rules required for accurate translation between LOE and LSE. Similar projects to LSEAvatar have been proposed in other Spanish speaking countries; for instance, in Mexico [9] or Ecuador [10]; however, they work with the sign language of those countries that differs from LSE. Additionally, services such as SVisual [11], available in police stations, connect LSE users with interpreters. However, such systems are not available in most everyday situations, and usually, deaf people have to pay to access this kind of service. Therefore, our project aims to significantly expand communication opportunities for deaf individuals across a wide range of environments and scenarios, ultimately providing an inclusive and accessible tool.

2. Architecture

The avatar that will be built in the LSEAvatar project consists of three modules depicted in Figure 1. All the models, datasets and code associated with the project will be publicly released with an open-source license.

The first module of the avatar is responsible for transcribing LOE audio in LOE text. To achieve this, pre-trained multilingual speech-to-text models such as Whisper [12] or Seamless [13] will be used. These models will be evaluated by taking into account the dialect of the speakers and their speed rate.

The second module focuses on translating LOE text into LSE glosses. This can be seen as a machine translation problem, for which the most successful approach to date is sequence-to-sequence models based on transformers [14]. In our case, we will fine-tune and compare several of these models, using different open-source language models in Spanish (such as Bertin [15] or Maria [16]) and multilingual models, such as mt5 [17], as starting points.

The third module will be responsible for extracting the necessary movements to sign the different LSE signs from the glosses. To carry out this development, we will use Spanish Sign Language Dictionaries, which contain videos of more than 10,000 signs performed by deaf professionals specialized in LSE. From these videos, we will extract the necessary signing movements using open-source libraries such as MediaPipe [18] or OpenPose [19], which enable tracking of key points on the hands, arms, face, and body. For words without an associated sign, the fingerspelling alphabet will be used.

Finally, to implement the avatar, we will use Blender [20], a free and open-source platform dedicated to modelling, animation, and the creation of three-dimensional graphics.

3. Results

In this section, we present the results obtained in the construction of each module of the project.

3.1. Speech to text

The first module is in charge of transcribing LOE audio in LOE text; therefore, we have analysed several pretrained automatic speech recognition (ASR) models in the open-source COSER corpus (that stands for *Corpus Oral y Sonoro del Español Rural*, in English Audible Corpus of Rural Spanish) [21], which captures the varieties of spoken European Spanish — the results of our analysis were presented in [22]. In particular, we considered 7 models: 6 Whisper-based [12], tiny, base, small, medium, large-v2, and large-v3; and 1 Seamless [13] model, SeamlessM4T v2 large model.

The mean and standard deviation of the performance of each ASR model is shown in Figure 2. As we can see in those results, the mean performance in terms of Word Error Rate (WER) of the models ranges from 0.81 in the case of the Whisper tiny model, to 0.292 in the case of the Whisper large v3 model (the lower, the better). Moreover, we can notice that, as expected, increasing the size of the Whisper model reduces the errors produced by the model. However, there is not a significant difference between the large v2, large v3 and medium versions of Whisper; hence, in this context, the version trained with more data does not provide a significant benefit. Finally, the performance of the Seamless model is only better than the tiny version of Whisper; therefore, this model does not seem a suitable

Table 1
Inference times of the ASR models.

Model	Time (secs)
Whisper tiny	4.75
Whisper base	5.32
Whisper small	8.69
Whisper medium	15.25
Whisper large-v2 and -v3	23.89
Seamless	4.36

alternative to the family of Whisper models.

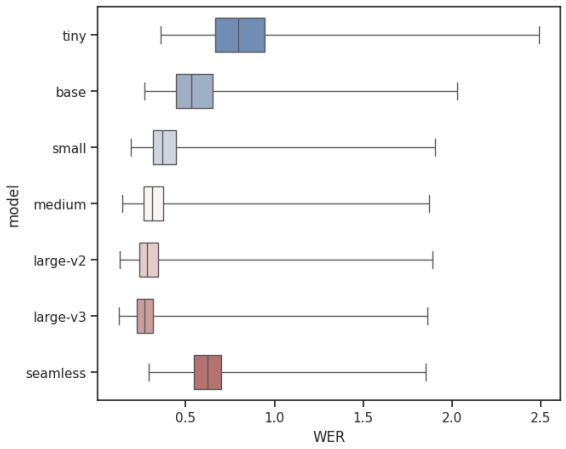


Figure 2: Box and whisker graph that represents how each model behaves for the analyzed audios.

Additionally, we have studied how much time it takes for each ASR model to process 1 minute of audio using a GPU NVIDIA GeForce RTX 3080, see Table 1 — note that the two large versions of Whisper take the same time. In the case of the Whisper models, the bigger the model, the slower; namely, the tiny version took approximately 4.75 seconds to process the audio, but the large models took almost 24 seconds. It is worth noticing that the Seamless model is the fastest of the analysed ASR systems, even faster than the Whisper tiny model, but as we previously mentioned, its performance is not on par with the bigger models of the Whisper family.

From these results, we can conclude that the large v3 version of Whisper produces the most accurate transcriptions; however, it is considerably slower than its smaller counterparts. In the context of building the transcriptions from a recorded video, processing time is not usually an issue, since the automatic transcription process can be run in the background, and after it finishes, it can be fed to the following module of our architecture. Nevertheless, large models might require special hardware to run in a reasonable time and are not suitable to be used for real-time processing; in such cases, the medium version of Whisper provides a good trade-off between accuracy in the transcription and inference speed. For this project, we decided to implement the large v3 version of Whisper since our system will initially work in an offline manner, and will not require real-time processing. Therefore, it is better to obtain more accurate transcriptions even if they take more time to process.

Table 2
Hyperparameters of models trained for translating LOE text to LSE glosses.

Model	Learning rate	Weight decay	Batch size	Epochs
MBart Large 50	5.6e-5	0.01	3	3
Marian MT	5.6e-5	0.01	4	4
T5-small	5.6e-4	0.05	4	6

Table 3
Performance of models trained for translating LOE text to LSE glosses.

Model	BLEU	BLEU-3	BLEU-4	Rouge-L
MBart Large 50	70.34	64.92	56.47	88.89
Marian MT	67.25	60.34	51.38	87.68
T5-small	36.90	45.21	34.21	68.75
Baseline [23]	57.61	9.98	4.98	46.52

Table 4
Number of videos downloaded from each LSE dictionary.

	ARASAAC	Sématos	Spreadthesign
Videos	4,139	5,911	13,439
Unique signs	1,617	3,328	9,423

3.2. LOE to glosses

The second module is in charge of translating LOE text to LSE glosses. We have approached this task as a translation problem and trained several sequence-to-sequence models. In order to train those models, we have used the synLSE dataset presented in [23]. The synLSE dataset is, as far as we are aware, the only dataset devoted to translate Spanish text to glosses. The synLSE dataset was a synthetically created corpus, and this might introduce potential limitations in terms of linguistic variability, naturalness, and generalization to real-world scenarios; hence, further research is necessary to incorporate real-world data.

Using the synLSE dataset, we fine-tuned three multilingual models (MBart Large 50 [24], Marian MT [25] and T5-Small [26]) on Google Colab with the by-default hyperparameters shown in Table 2 and using the functionality provided by the Huggingface library [27]. Finally, we have evaluated the models using both BLEU, BLEU-3, BLEU-4 and Rouge-L [28].

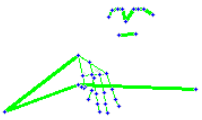
The performance of the different models on the test set of the synLSE dataset is presented in Table 3. From those results, we can conclude that the best model was obtained with the MBart architecture, with a BLEU-4 of 56.47 and a ROUGE-L of 88.89, considerably improving the baseline results presented in [23].

3.3. Avatar

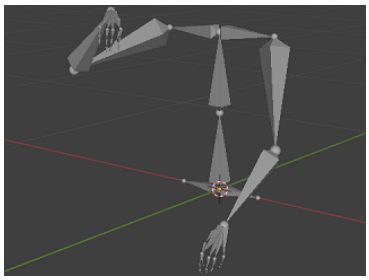
The last module of the project is the construction of an avatar that signs LSE glosses. The construction of this module is still ongoing, and the preliminary steps are presented here.

First of all, we have built a dataset of LSE glosses captured from three different sources: ARASAAC [29], Sématos [30] and Spreadthesign [31]. A total of 23,488 videos were downloaded, 14,368 of them from unique signs — the number of videos and unique signs from each site can be seen in Table 4.

(a) Frame of a person signing the letter m from the Spreadthesign dataset.



(b) Skeleton extracted from the frame in Figure 3a.



(c) 3D armature obtained from the skeleton in Figure 3b. This armature is generated in Blender.

Figure 3: Conversion of a person's pose into a Blender armature.

The next step is the extraction of the movements from each video. We have studied several 3D Human Pose Estimation models for this step. To the best of our knowledge, there are two different approaches. In the first approach, models directly estimate 3D locations of human joints; whereas, in the second approach, a model first estimates 2D locations of human joints, and then another model lifts the estimations to 3D. Some estimators we have tested are: for direct 3D estimations Mediapipe [18]; for 2D estimations OpenPose [19] and AlphaPose [32]; and for lifting 2D estimations MotionBert [33] and the method presented in [34]. Figure 3 illustrates an example obtained with these approximations in which a person's pose from a video frame is converted into a Blender armature by extracting the locations of the human's joints. Furthermore, the reader can watch a video of the armature signing "LSEAvatar" in LSE using fingerspelling at the link https://www.youtube.com/shorts/BB-KhEqMAu4.

Several tasks remain as future work to advance the development of the avatar. These include the seamless integration of multiple video segments to generate coherent and natural sign language sentences, the incorporation of a full-body model into the existing 3D armature, and the comprehensive validation of the constructed avatar. The evaluation of such technologies is typically carried out manually by experts, who assess dimensions such as grammatical accuracy in sign language, naturalness of expression, readability, and cultural appropriateness [35]. Additionally, some prior studies have employed back-translation techniques as a complementary evaluation method [36]. In this work, we plan to adopt both expert-based and back-translation approaches to assess the performance of LSEAvatar.

4. Social Impact

LSEAvatar aims to benefit all individuals who use Spanish Sign Language (LSE) as their primary means of communication, which amounts to approximately 70,000 people [3]. Additionally, when a deaf person requires simultaneous translation through an interpreter, the deaf individual currently bears both the cost and the responsibility of securing the interpreter. The tool developed as a product of the LSEAvatar project will benefit deaf individuals, who will gain direct access to relevant information, as well as public administrations and businesses, which will be able to disseminate information to a larger audience.

This tool must be validated by LSE experts for it to be helpful to the deaf community. To this end, we are collaborating with the Association of the Deaf of La Rioja. Specifically, between three and eight individuals from the association will participate in the validation process. Members of this association will be the first to test the tool and, consequently, benefit from its use.

As the project expands and more users rely on the avatar for translation, a scalable computing infrastructure will be required to handle the workload. This entails leveraging cloud computing technologies that allow for vertical and horizontal scaling as needed. Scalability also involves continuously improving and updating the avatar's various components to enhance translation accuracy and fluency. Finally, as the project grows, it will be crucial to ensure that the avatar is compatible with a wide range of devices and platforms, such as mobile applications, websites, and smart devices, necessitating a flexible and adaptive development approach.

Furthermore, this project has the potential to be adapted to other sign languages, thereby increasing its reach and impact. It is important to consider that, just as there are numerous spoken languages, the same applies to sign languages, estimated over 200 different sign languages worldwide. In fact, the sign languages used in Spanish-speaking countries vary, meaning that the avatar developed for Spain cannot be directly used elsewhere. However, the avatar will have a modular design, requiring only the development of a translation model from the spoken language to the corresponding sign language, along with the provision of a dataset of sign language videos to adapt the avatar accordingly.

Funding institutions

This work was partially supported by INDRA through the call "*Convocatoria de Ayudas de Proyectos de Investigación en Tecnologías Accesibles 2024*" and by the Government of La Rioja through Proyecto INICIA 2023/01 and AFIANZA 2024/01.

Research groups

The development of LSEAvatar is conducted by members of two groups of University of La Rioja: the *Grupo de Informática de la Universidad de La Rioja* (https://investigacion.unirioja.es/grupos/45/detalle), and the *Grupo de Lingüística Aplicada de la Universidad de La Rioja* (https://investigacion.unirioja.es/grupos/4/detalle).

Acknowledgments

We are grateful to the *Asociación de Personas Sordas de La Rioja* for their help in the development of this project. We thank M. Ivashechkin for his help with the task of 3D Human Pose Estimation.

References

[1] World Federation of the Deaf, Our work, 2024. URL: https://wfdeaf.org/our-work/.

[2] INE, Utilización de la lengua de signos por sexo y edad. Población de 6 y más años con discapacidad de audición, 2024. URL: https://www.ine.es/, accessed: 2025-03-13.

[3] Confederación Estatal de Personas Sordas (CNSE), Personas sordas, 2024. URL: https://www.cnse.es/index.php/personas-sordas, accessed: 2024-03-20.

[4] I. Rodríguez-Moreno, J. M. Martínez-Otzeta, B. Sierra, A Hierarchical Approach for Spanish Sign Language Recognition: From Weak Classification to Robust Recognition System, in: Intelligent Systems and Applications, Elsevier, 2023, pp. 37–53.

[5] A. Nuñez-Marcos, et al., A survey on Sign Language machine translation, Expert Systems with Applications 213 (2023) 118993. doi:10.1016/j.eswa.2023.118993.

[6] Google, Live Transcribe & Notification, 2024. URL: https://play.google.com/store/apps/details?id=com.google.audio.hearing.visualization.accessibility.scribe&hl=en_US, accessed: 2025-03-14.

[7] F. Morillas-Espejo, E. Martinez-Martin, Sign4all: A Low-Cost Application for Deaf People Communication, IEEE Access 11 (2024) 98776–98786. URL: https://ieeexplore.ieee.org/document/10242052.

[8] F. Morillas-Espejo, E. Martinez-Martin, A virtual avatar for sign language signing, in: International Conference on Soft Computing Models in Industrial and Environmental Applications, Springer, 2024, pp. 58–67.

[9] B. Martinez-Seis, O. Pichardo-Lagunas, E. Hernández-Morales, O. Rivera-Rodríguez, S. Miranda, Automatic translation of sentences to mexican sign language: Rule-based machine translation and animation synthesis in avatar, Computación y Sistemas 29 (2025) 145–155.

[10] C. Salamea-Palacios, K. A. Salcedo, M. Peralta-Marin, E. J. Sacoto-Cabrera, Prototype of a text to ecuadorian sign language translator using a 3d virtual avatar, in: 2024 IEEE Colombian Conference on Communications and Computing (COLCOM), IEEE, 2024, pp. 1–6.

[11] Servicio SVisual en las comisarías de la Policía Nacional 091, 2024. URL: https://www.svisual.org/, accessed: 2025-05-26.

[12] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever, Robust speech recognition via large-scale weak supervision, in: International Conference on Machine Learning, PMLR, 2023, pp. 28492–28518.

[13] L. Barrault, Y.-A. Chung, M. C. Meglioli, D. Dale, N. Dong, P.-A. Duquenne, H. Elsahar, H. Gong, K. Heffernan, J. Hoffman, et al., Seamlessm4t-massively multilingual & multimodal machine translation, arXiv preprint arXiv:2308.11596 (2023).

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017) 1–11.

[15] J. De la Rosa, E. G. Ponferrada, P. Villegas, P. G. d. P. Salas, M. Romero, M. Grandury, Bertin: Efficient pre-training of a spanish language model using perplexity sampling, arXiv preprint arXiv:2207.06814 (2022).

[16] A. Gutiérrez-Fandiño, J. Armengol-Estapé, M. Pàmies, J. Llop-Palao, J. Silveira-Ocampo, C. P. Carrino, A. Gonzalez-Agirre, C. Armentano-Oller, C. Rodriguez-Penagos, M. Villegas, Maria: Spanish language models, arXiv preprint arXiv:2107.07253 (2021).

[17] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, C. Raffel, mt5: A massively multilingual pre-trained text-to-text transformer, arXiv preprint arXiv:2010.11934 (2020).

[18] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, et al., Mediapipe: A framework for building perception pipelines, arXiv preprint arXiv:1906.08172 (2019).

[19] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, Y. A. Sheikh, Openpose: Realtime multi-person 2d pose estimation using part affinity fields, IEEE Transactions on Pattern Analysis and Machine Intelligence (2019).

[20] R. Hess, Blender foundations: The essential guide to learning blender 2.5, Routledge, 2013.

[21] I. Fernández-Ordóñez, COSER. Corpus Oral y Sonoro del Español Rural, 2005.

[22] M. San Martín, J. Heras, G. Mata, S. Gómez, Is ASR the right tool for the construction of Spoken Corpus Linguistics in European Spanish?, Procesamiento del lenguaje natural 73 (2024) 165–176.

[23] M. Perea-Trigo, C. Botella-López, M. Á. Martínez-del Amor, J. A. Álvarez-García, L. M. Soria-Morillo, J. J. Vegas-Olmos, Synthetic corpus generation for deep learning-based translation of spanish sign language, Sensors 24 (2024) 1472.

[24] Y. Tang, C. Tran, X. Li, P.-J. Chen, N. Goyal, V. Chaudhary, J. Gu, A. Fan, Multilingual translation with extensible multilingual pretraining and finetuning, arXiv preprint arXiv:2008.00401 (2020).

[25] M. Junczys-Dowmunt, R. Grundkiewicz, T. Dwojak, H. Hoang, K. Heafield, T. Neckermann, F. Seide, U. Germann, A. Fikri Aji, N. Bogoychev, A. F. T. Martins, A. Birch, Marian: Fast neural machine translation in C++, in: Proceedings of ACL 2018, System Demonstrations, Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 116–121. URL: http://www.aclweb.org/anthology/P18-4020.

[26] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, Journal of Machine Learning Research 21 (2020) 1–67. URL: http://jmlr.org/papers/v21/20-074.html.

[27] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Huggingface's transformers: State-of-the-art natural language processing, arXiv preprint arXiv:1910.03771 (2019).

[28] E. Chatzikoumi, How to evaluate machine translation: A review of automated and human metrics, Natural Language Engineering 26 (2020) 137–161.

[29] ARASAAC, Last accessed on June 2025. URL: https://arasaac.org/.

[30] Sématos, Last accessed on June 2025. URL: https://www.sematos.eu/lse.html.

[31] Spreadthesign, Last accessed on June 2025. URL: https://spreadthesign.com/es.es/search/.

[32] H.-S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y.-L. Li, C. Lu, AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time, IEEE Transactions on Pattern Analysis & Machine Intelligence 45 (2023) 7157–7173. URL: https://doi.ieeecomputersociety.org/10.1109/TPAMI.2022.3222784. doi:10.1109/TPAMI.2022.3222784.

[33] W. Zhu, X. Ma, Z. Liu, L. Liu, W. Wu, Y. Wang, MotionBERT: A Unified Perspective on Learning Human Motion Representations , in: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE Computer Society, Los Alamitos, CA, USA, 2023, pp. 15039–15053. URL: https://doi.ieeecomputersociety.org/10.1109/ICCV51070.2023.01385. doi:10.1109/ICCV51070.2023.01385.

[34] M. Ivashechkin, O. Mendez, R. Bowden, Improving 3d pose estimation for sign language, arXiv preprint arXiv:2308.09525 (2023).

[35] Z. Yuan, Z. Ruiquan, Y. Dengfeng, C. Yidong, Translation quality evaluation of sign language avatar, in: Proceedings of the 23rd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations), 2024, pp. 405–415.

[36] R. Zuo, F. Wei, Z. Chen, B. Mak, J. Yang, X. Tong, A simple baseline for spoken language to sign language translation with 3d avatars, in: European Conference on Computer Vision, Springer, 2024, pp. 36–54.