# Matignon-LSF: a Large Corpus of Interpreted French Sign Language

Julie Halbout, Diandra Fabre, Yanis Ouakrim, Julie Lascar, Annelies Braffort, Michèle Gouiffès, Denis Beautemps

# Matignon-LSF: a Large Corpus of Interpreted French Sign Language

**Julie Halbout*[1], Diandra Fabre*[2], Yanis Ouakrim*[1,2], Julie Lascar*[1]**
**Annelies Braffort[1], Michèle Gouiffès[1], Denis Beautemps[2]**
[1]Univ. Paris-Saclay, CNRS, LISN, Orsay, France
[2]Univ. Grenoble Alpes, CNRS, GIPSA-Lab, Grenoble, France
[1]firstname.lastname@lisn.upsaclay.fr,
[2]firstname.lastname@gipsa-lab.grenoble-inp.fr

**Abstract**

In this paper we present Matignon-LSF, the first dataset of interpreted French Sign Language (LSF) and one of the largest LSF dataset available for research to date. This is a dataset of live interpreted LSF during public speeches by the French government. The dataset comprises 39 hours of LSF videos with French language audio and corresponding subtitles. In addition to this data, we offer pre-computed video features (I3D). We provide a detailed analysis of the proposed dataset as well as some experimental results to demonstrate the interest of this novel dataset.

**Keywords:** French Sign Language, LSF, dataset, interpretation, alignment

## 1. Introduction

Automatic processing of sign languages (SL) is an expanding field, but unfortunately the vast majority of these languages are still poorly endowed in terms of corpora available for research. This is particularly the case for French Sign Language (LSF). One potential source of SL data is television (Koller et al., 2015; Albanie et al., 2021), where the number of interpreted programs has increased in recent years. However, the access to this data is generally not easy for research purposes, due to rights or technical problems. In France, weekly Council of Ministers debriefings yield open-access videos which are systematically interpreted in LSF. We have taken advantage of this opportunity to compile a new dataset called Matignon-LSF[1] (fig. 1), which is presented in this paper.

The primary language modality of the TV programs is speech. Speech may be subtitled, sometimes in real time, either automatically with all the potential errors that this entails, or in a live subtitling studio with time and format constraints. Speech may also be interpreted in SL, sometimes in post-production, which enables the SL version to be prepared and corrected, or sometimes in real time. In this last scenario, several phenomena occur. Usual practice in interpreting is for the professional to interpret into their native language. The situation is different in the case of SL interpreting because it is necessary for the interpreter to hear speech. Therefore, unless the interpreter is a CODA (child of deaf adult), he/she interprets into a second language. In addition, there is some evidence of differences between the output of hearing and deaf interpreters

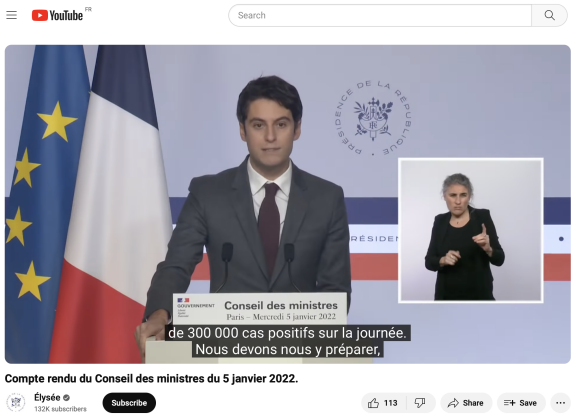


Figure 1: Screenshot from a video in the Matignon-LSF dataset, showing debriefings from the French government's Council of Ministers.

(Stone and Russell, 2011). Furthermore, due to strong time constraints, SL during real-time interpretation tends to closely follow the grammatical structure of the spoken language, with evidences that differences in forms of language are reduced in interpreted content (Dayter, 2019). The interpreters may choose not to convey information from the audio stream that they consider to be redundant to the visual stream of the footage. Fluent signers can generally tell the difference between interpreted and non-interpreted SL, as well as signing by native deaf signers and non-native or non-deaf signers.

It is worth emphasizing that, due to the interpretation process, the source language can interfere in the signing. Thus interpreted SL can be different from original SL (i.e. directly produced by signers). However, there is little work on describing or quantifying these differences.

Having said that, this kind of dataset may be very

*These authors contributed equally to this work and none of the authors are Deaf

[1]*Matignon* refers to the official residence of the French Prime Minister, and in extends to the french government.



*Proceedings of the 11th Workshop on the Representation and Processing of Sign Languages*, pages 95–101

useful in automatic processing because it provides more SL data, even if it is task specific. In our case, it also has the advantage of being open-data.

In this paper, after a brief overview of the corpora currently available (section 2), particularly in LSF, section 3 presents the Matignon-LSF dataset, the collection and processing of the data, and section 5 discusses the perspectives opened by this new dataset.

## 2. Related Work

As part of the recent Easier European project, an overview of existing datasets for the European SLs was drawn up (Kopf et al., 2023). These datasets were divided in two categories: linguistic corpora and broadcast data. The former offer high-quality data with rich transcriptions and annotations, while the latter are available in large quantities. Since the publication of this report, other datasets have been released, such as BSL-1K (Albanie et al., 2020) and more recently BOBSL in British Sign Language (BSL) (Albanie et al., 2021), which represents a change of scale in terms of dataset, providing researchers with over 1,200 hours of sign language interpreted from BBC broadcasts. In a similar vein, the American Sign Language YouTube-ASL dataset (Uthus et al., 2024) totals almost 1,000 hours of videos from the web. Also in ASL, the How2Sign corpus, published in 2023, is of particular interest, as it is the largest laboratory corpus of original (non interpreted or translated) SL. This has already been the subject of several works (Duarte et al., 2021).

LSF has been the subject of several corpora collections over the last 10 years (Braffort, 2022). Most of these LSF corpora have been compiled in laboratories mainly for linguistic research works, and have two main shortcomings: fully annotated datasets like *Rosetta* and *40 brèves* are very small, containing less than 4 hours of data and larger datasets, such as Creagest (Balvet et al., 2010), are only partially annotated. The DictaSign dataset, consisting of 8-hour dialogues (Belissen et al., 2020), is currently partially annotated. Nevertheless, it remains valuable for recognizing signs in context, including lexical (Ouakrim et al., 2023) and non-lexical instances (Belissen et al., 2020).

Recently, two LSF datasets have been made available to overcome these problems: Mediapi-Skel (Bull et al., 2019) and Mediapi-RGB (Ouakrim et al., 2024). The last one comprises 86 hours of videos in LSF produced by deaf reporters or presenters from the bilingual online medium Média'Pi! with French subtitles produced by deaf translators. The subtitles are well-aligned with LSF videos, and the dataset has been prepared for processing (Ouakrim et al., 2024). These two corpora are in a LSF-to-French modality because subtitles were produced accordingly to the signing (and not the other way around) and are perfectly aligned. These two corpora are much larger than the previous ones in LSF, except for the non completely annotated Creagest corpus.

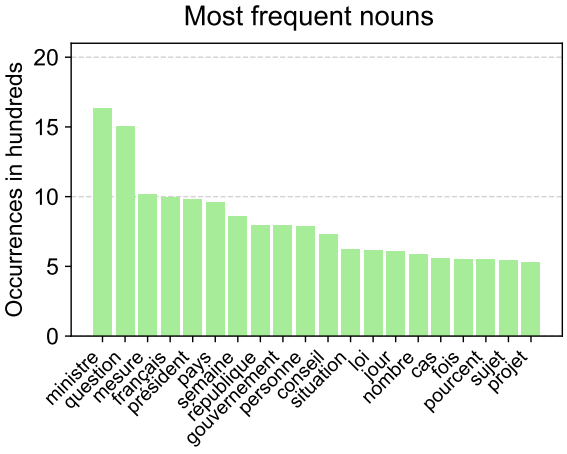


Figure 2: 20 most frequent nouns in the subtitles of Matignon-LSF.

Due to the economic model of this medium, videos are unavailable for Mediapi-Skel and only partially available for Mediapi-RGB, which may be a limitation for researchers wishing to use features other than those pre-extracted by the authors (body pose, I3D, etc.).

Thus, our aim is to collect a new LSF dataset that is both large and open. We are therefore interested in interpretation data from French broadcast and created the Matignon-LSF dataset detailed in the following sections.

## 3. Dataset overview

French government's Council of Ministers debriefings take place once a week at l'Elysée. They are filmed, subtitled and, since July 2020, interpreted in LSF. The Matignon-LSF dataset is based on the LSF interpretations and subtitles of theses debriefings. We do not have further information yet regarding the work process of the interpreters, but they probably don't have much material to prepare their interpretation. To date, it includes 67 debriefing videos. Figure 2 shows the 20 most frequent nouns of the dataset, demonstrating that the content of the speech is strongly related to French politics (top five words: *minister*, *question*, *measure*, *french* and *president*).

59 videos consist of the government spokesperson's speech (which varies from 4 minutes to 20 minutes, with an average of about 12 minutes), followed by a question-and-answer session with journalists. This part can vary depending on the

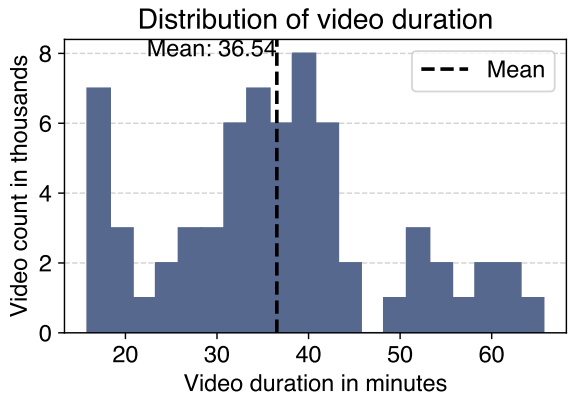


Figure 3: Video's duration distribution.

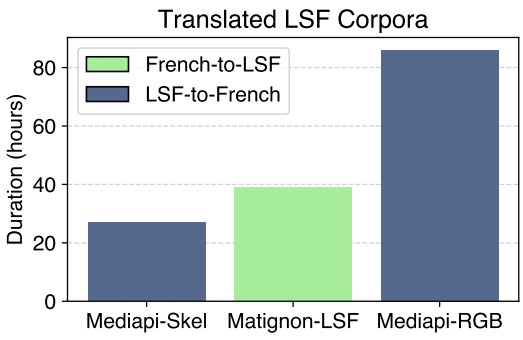


Figure 4: Duration of translated LSF Corpora. Matignon-LSF is the second largest translated LSF Corpora after Mediapi-RGB and the largest interpreted LSF corpora.

topics and the number of journalists in the press room (from 8 minutes to almost an hour, with an average of 23 minutes). In five other videos, ministers are invited to present their points after the spokesperson's speech, and they are asked questions in addition to the spokesperson. In the three remaining videos, the press conference is held without a spokesperson, and the ministers deliver their speeches directly, with a shorter question-and-answer session. The 67 delivered videos have a total duration of 39 hours, with an average duration of 36 minutes. The distribution of video duration is shown in Figure 3.

The subtitles (written French) in the dataset is composed of a total of 447k tokens for a total vocabulary size of 10k[2]. From the subtitles, we extracted 18k sentences, as described in section 4.3. Matignon-LSF features 15 signers.

The characteristics of the dataset are summarized in the table 1.

| | |
|---|---|
| **Total duration (h)** | 39 |
| **#videos** | 67 |
| **#subtitles** | 51131 |
| **#sentences** | 18000 |
| **#french words vocab.** | 10000 |
| **#signers** | 15 |
| **#speakers** | 3* |
| **Video resolution (px)** | $494 \times 494$ |
| **Frame rate (fps)** | 30 |

Table 1: **Dataset overview.** *journalists and ministers not included.

To date, corpus Matignon-LSF lies between Mediapi-Skel and Mediapi-RGB in terms of size (fig. 4).

[2]we used SpaCy tokenizer https://spacy.io/

## 4. Data collection and processing

This section details the construction of the Matignon-LSF dataset. We present the raw data and the processing carried out to provide the dataset. The diverse processes are documented in a GitHub repository, organized as a toolbox to enable reproduction and expansion of the corpus, as new press release takes place once a week.

### 4.1. Collecting the SL videos and subtitles

Each week, the debriefing is filmed and uploaded on Youtube and/or Dailymotion and comes with a corresponding set of written French subtitles aligned with the audio. Original videos have a resolution of 1080 px and a frame rate of 30 fps.

Using the PyTube Python library, we downloaded all videos issued between December 2020 and December 2023 along with their associated audio track. We then used the YouTube Transcript Python Api to download the subtitles, and keep only manually written subtitles, setting aside videos that only have generated subtitles. Obtained JSON files are then converted to the VTT subtitle format. Next, using OpenCV, we crop the videos so as to retain only the square containing the LSF interpreter.

After the above steps, we obtain $494 \times 494$ px LSF videos with associated French audio and subtitles.

### 4.2. Processing the videos

Skeleton keypoints, such as those provided by OpenPose (Cao et al., 2018) and Mediapipe Holistic (Lugaresi et al., 2019), are essential inputs for various automated sign language processing tasks. These tasks include cropping of hands or faces (Huang et al., 1994), generating sign language (Ventura et al., 2020), and improving recognition methods (Belissen et al., 2020).

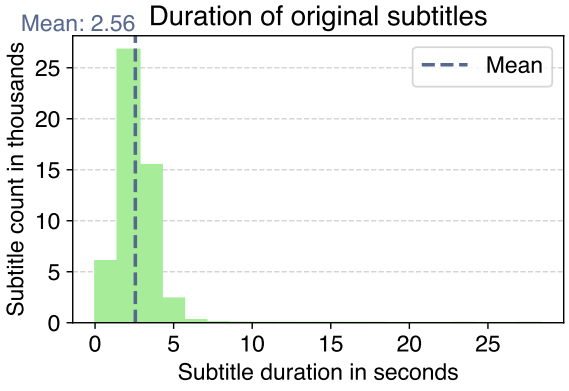


Figure 5: Distribution of subtitle duration before sentences extraction.

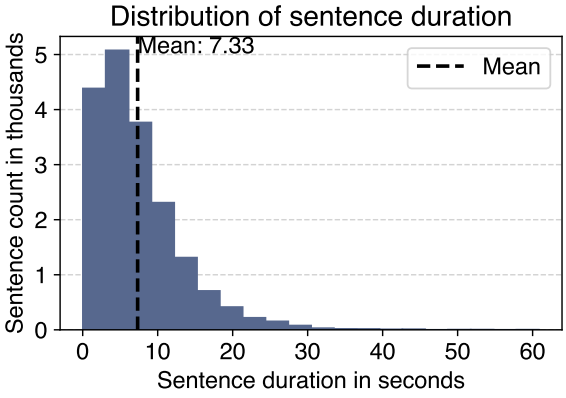


Figure 6: Distribution of sentence duration after sentences extraction.

Other automatic sign language processing methods (Tarrés et al., 2023; Renz et al., 2021) rely on features extracted from sign language videos by the I3D model (Carreira and Zisserman, 2017). We used this architecture to extract features from our videos. Specifically, we have used the fine-tuned model provided by Varol et al. (2021).

### 4.3. Processing the subtitles

As subtitles are constrained by length for display reason, they do not necessarily form sentences. However, the translation tasks often operate at the sentence level.

To address this, we generate a sentence-level segmentation from the subtitles. We adopt the same approach as Albanie et al. (2021) to build our sentence-segmented subtitle files. We split subtitles on sentence boundary punctuation. When a sentence spans multiple subtitles, it is easy to extract the sentence by concatenation. It is more complicated when multiple sentences fall in one subtitle. As the method used by Albanie et al. (2021), to preserve the alignment, we calculate the duration of a character (based on the subtitle's characters length). We can use this information to associate a duration to each sentence within the subtitle. Then, we can calculate the new subtitle's timestamps on this basis. The disparity of the subtitle's duration between the original subtitles and the sentence-segmented subtitles is illustrated in Figures 5 and 6. The average time thus increases from 2.56 to 7.33 seconds.

The corpus will be soon deposited on the Ortolang platform and will be regularly updated over time. We estimate that it should be able to increase by around 13h per year.

## 5. Perspectives

The Matignon-LSF corpus has a number of advantages that can be exploited to address various computer vision and natural language processing tasks.

**Alignment.** At this stage, the French subtitles and LSF of Matignon-LSF are not yet aligned as can be seen in Fig 7. This example shows two consecutive sentences. "Un cap pour contrôler l'épidémie. Un cap pour relancer notre pays." (*A direction to control the epidemic. A direction to relaunch our country.*). We observed that the length of the two signed sentences (4.64 seconds) is longer than that of the two spoken sentences (3.9 seconds). Therefore, a manual shift of the speech subtitles is not enough to fit the data: the GT and Sub alignments would start at the same time, but end differently.

Whatever the type of language (spoken, written or signed), machine translation methods require prior alignment between the source and the target languages. In order to use this dataset for translation tasks, it is necessary to be able to associate an extract of LSF with its corresponding French subtitles. The Matignon-LSF dataset contains a complete translation for each of the 67 videos. However, providing 35-minute video sequences (±52,500 frames) and their associated translations to a translation model would be very costly. It would therefore be necessary to divide these videos into sub-extracts.

State-of-the-art methods mostly rely on sentence segmentation. Hence, videos and text are split into sentence-like units, with an association between text and SL: for each SL sentence, the text corresponding to the translation is given. However, producing such an SL sentence/text alignment from an interpreted SL dataset is a real challenge: the text is aligned with the audio, whereas SL interpretation is performed with a latency that varies

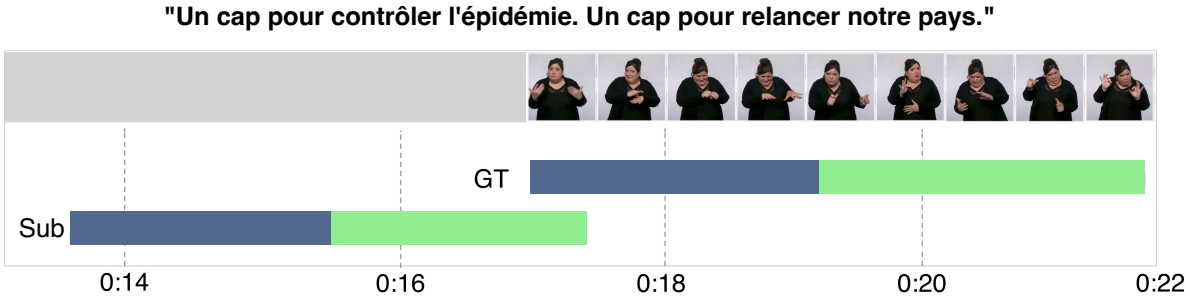


Figure 7: Demonstrating the alignment challenge in Matignon-LSF. The GT line corresponds to a manual alignment (or Ground Truth) annotated for this specific figure while the Sub line corresponds to the subtitles as provided with Matignon-LSF. Blue block corresponds to one sentence while green block corresponds to another sentence.

in time and from one interpreter to another. Thus the very first task to be carried out on this dataset should be to align the subtitles with the LSF content. Manual alignment requires a considerable time commitment as explained in (Bull et al., 2020): It takes an expert fluent in sign language approximately 10-15 hours to synchronize subtitles with one hour of continuous sign language video. Automatic alignment methods as the one used for the BOBSL dataset (Bull et al., 2021) could be a solution but might need some fine-tuning for LSF.

**Sign Language modeling.** The Matignon-LSF dataset can be used as it is, with no need for prior alignment, for sign language modeling and can be used to train unsupervised language models on LSF such as SignBERT (Hu et al., 2021).

**Sign Recognition.** With the help of a method like Lascar et al. (2024)'s automatic annotation process currently under development, we could perform automatic sign recognition and classification. This would provide information on the number of lexical signs in our dataset. Sign classification is also a step towards aligning our dataset between SL and the subtitles. However, one should note that the sign interpreters produce an interpretation of the speech that appears in the subtitles, as opposed to a transcription. This means that words in the subtitles may not correspond directly to individual signs produced by the interpreters, and vice versa. There may also be discrepancies between the audio and the subtitle text.

**Sign Language Translation.** Once aligned, the Matignon-LSF dataset could be used to train machine translation models for a wide variety of modalities: LSF to French text, LSF to Speech, and vice-versa (Ventura et al., 2020; Müller et al., 2023; Ouakrim et al., 2024).

**Studying interpreted LSF** As the first interpreted LSF dataset of this scale, Matignon-LSF can be used to study the specificity of interpreted LSF in comparison with the original LSF that can be observed in other corpora. For example, the work of (Belissen et al., 2020) could be used to quantify the distribution of sign types in this dataset.

## 6. Conclusion

In this paper, we presented Matignon-LSF, a new dataset completely open to both research and private use. We gave an overview of the dataset and then presented the processing steps we applied for the collection and preparation.

The scripts we developed are publicly available so that they may be used to extend the dataset as new videos are produced and published every week. We also aim at adding other videos such as President or Prime Minister solo intervention. The corpus itself will be soon made available on the Ortolang platform.

This dataset is the first dataset of interpreted LSF, also usable outside public research. Future work should focus on aligning this dataset, in particular to facilitate the suggested perspectives.

## Acknowledgements

This work has been partially funded by the Bpifrance investment "Structuring Projects for Competitiveness" (PSPC), as part of the Serveur Gestuel project.

## Authors details

None of the authors are deaf. A deaf colleague, specialist in motion capture and virtual signer animation, belongs to our team but didn't participate to this project. Moreover, we often collaborate with the Deaf community.

## Bibliographical References


S. Albanie, G. Varol, L. Momeni, T. Afouras, Joon S. Chung, N. Fox, and A. Zisserman. 2020. Bsl-1k: Scaling up co-articulated sign language recognition using mouthing cues. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 35–53. Springer.

S. Albanie, G. Varol, L. Momeni, H. Bull, T. Afouras, H. Chowdhury, N. Fox, B. Woll, R. Cooper, A. McParland, et al. 2021. BOBSL: BBC-Oxford British Sign Language Dataset. In *ArXiv preprint*.

A. Balvet, C. Courtin, D. Boutet, C. Cuxac, I. Fusellier-Souza, B. Garcia, M.-T. L'Huillier, and M. A. Sallandre. 2010. The creagest project: a digitized and annotated corpus for french sign language (lsf) and natural gestural languages. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 469–475.

V. Belissen, A. Braffort, and M. Gouiffès. 2020. Experimenting the automatic recognition of non-conventionalized units in sign language. *Algorithms*, 13(12):310–336.

A. Braffort. 2022. Langue des signes française: Etat des lieux des ressources linguistiques et des traitements automatiques. In *Journées Jointes des Groupements de Recherche Linguistique Informatique, Formelle et de Terrain (LIFT) et Traitement Automatique des Langues (TAL)*, pages 131–138. CNRS.

H. Bull, T. Afouras, G. Varol, S. Albanie, L. Momeni, and A. Zisserman. 2021. Aligning subtitles in sign language videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11552–11561.

H. Bull, A. Braffort, and M. Gouiffès. 2020. MEDIAPI-SKEL -A 2D-Skeleton Video Database of French Sign Language With Aligned French Subtitles. In *12th Conference on Language Resources and Evaluation (LREC 2020)*, pages 6063–6068, Marseille, France.

Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*.

J. Carreira and A. Zisserman. 2017. Quo vadis, action recognition? a new model and the kinetics dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6299–6308.

D. Dayter. 2019. *Collocations in non-interpreted and simultaneously interpreted English: a corpus study*. Routledge.

A. Duarte, S. Palaskar, L. Ventura, D. Ghadiyaram, K. DeHaan, F. Metze, J. Torres, and X. Giro-i Nieto. 2021. How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2735–2744.

H. Hu, W. Zhao, W. Zhou, and H. Li. 2021. Signbert+: Hand-model-aware self-supervised pre-training for sign language understanding. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11087–11096.

C. Huang, Joseph L. Mundy, and Charles A. Rothwell. 1994. Model supported exploitation: Quick look, detection and counting, and change detection. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 144–151.

O. Koller, J. Forster, and H. Ney. 2015. Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141:108–125.

M. Kopf, M. Schulder, and T. Hanke. 2023. The sign language dataset compendium.

J. Lascar, M. Gouiffès, A. Braffort, and C. Danet. 2024. Annotation of lsf subtitled videos without a pre-existing dictionary. In *Workshop on the Representation and Processing of Sign Languages at the International Conference on Language Resources and Evaluation (sign-lang@LREC)*.

C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann. 2019. Mediapipe: A framework for building perception pipelines. *CoRR*, abs/1906.08172.

M. Müller, M. Alikhani, E. Avramidis, R. Bowden, A. Braffort, N. Cihan Camgöz, S. Ebling, C. España-Bonet, A. Göhring, R. Grundkiewicz, M. Inan, Z. Jiang, O. Koller, A. Moryossef, A. Rios, D. Shterionov, S. Sidler-Miserez, K. Tissi, and D. Van Landuyt. 2023. Findings of the second WMT shared task on sign language translation (WMT-SLT23). In *Proceedings of the Eighth Conference on Machine Translation*, pages 68–94, Singapore. Association for Computational Linguistics.

Y. Ouakrim, D. Beautemps, M. Gouiffès, T. Hueber, F. Berthommier, and A. Braffort. 2023. A

multistream model for continuous recognition of lexical unit in french sign language. In *29° Colloque sur le traitement du signal et des images"*, 2023-1182, pages 461–464. GRETSI - Groupe de Recherche en Traitement du Signal et des Images.

Y. Ouakrim, H. Bull, M. Gouiffès, D. Beautemps, T. Hueber, and A. Braffort. 2024. Mediapi-RGB: Enabling technological breakthroughs in french sign language (LSF) research through an extensive Video-Text corpus. *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2:139–148.

K. Renz, N. C. Stache, S. Albanie, and G. Varol. 2021. Sign language segmentation with temporal convolutional networks. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2135–2139. IEEE.

C. Stone and D. Russell. 2011. Interpreting in international sign: decisions of deaf and non-deaf interpreters. In *Proceedings of World Association of Sign Language Interpreters Conference*.

L. Tarrés, G. I. Gállego, A. Duarte, J. Torres, and X. Giró-i Nieto. 2023. Sign language translation from instructional videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5624–5634.

D. Uthus, G. Tanzer, and M. Georg. 2024. Youtube-asl: A large-scale, open-domain american sign language-english parallel corpus. *Advances in Neural Information Processing Systems*, 36.

G. Varol, L. Momeni, S. Albanie, T. Afouras, and A. Zisserman. 2021. Read and attend: Temporal localisation in sign language videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16857–16866.

L. Ventura, A. Duarte, and X. Giró-i Nieto. 2020. Can everybody sign now? exploring sign language video generation from 2d poses. *arXiv preprint arXiv:2012.10941*.

## Language Resource References

Belissen, V. and Braffort, A. and Gouiffès, M. 2020. *Dicta-Sign-LSF-v2 corpus*. ISLRN 442-418-132-318-7.

Bull, H. and Braffort, A. and Gouiffès, M. 2019. *Mediapi-Skel corpus*. ISLRN 184-726-682-550-4.

Bull, H. and Ouakrim, Y and Lascar, J. and Braffort, A. and Gouiffès, M. 2024. *Mediapi-RGB corpus*. ISLRN 421-833-561-507-6.