Learning to Spot Signs from Named Entities. A study on French Sign Language.

Julie Halbout[1] Annelies Braffort[1] Michèle Gouiffès[1] Diandra Fabre[2] Julie Lascar[1]

[1]Univ. Paris-Saclay. CNRS. LISN, [2]Univ. Grenoble Alpes. CNRS. GIPSA-Lab
[1]91400, Orsay, France, [2]38000, Grenoble, France
[1]{julie.halbout, michele.gouiffes, annelies.braffort, julie.lascar}@lisn.upsaclay.fr
[2]diandra.fabre@gipsa-lab.grenoble-inp.fr

Abstract

French Sign Language (LSF) is a low-resourced language, with few available corpora, most of which being only partially annotated. Previous work on other sign languages has explored automatic sign annotation using subtitles as weak supervision, existing signaries, or mouthing cues. This paper focuses on the corpus Matignon-LSF, by first leveraging lexical token spotting then by studying Named Entities (locations, companies, persons). Accounting for the named entities enables the automatic detection of 30% to 100% more signs per class and improves the spotting of rare signs. In addition, this work provides insights into the signing of named entities and contributes resources for improving LSF-to-French translation models.

1. Introduction

Sign languages (SLs) are natural languages that rely on visual-gestural modalities rather than vocal or written forms. They are characterized by complex spatial and temporal structures, the use of co-occurring facial expressions and body movements, and non-monolinear syntax that differ significantly from vocal languages. French Sign Language (LSF) is considered as a low-resourced language (lack of data, technical tools, and linguistic resources). Moreover, only a few studies address LSF processing.

A recent study focused on a co-articulated sign spotting task, *i.e.* spotting of a specific sign, in a continuous LSF sequence (Lascar et al., 2024). This task has led to the creation of an initial signary (set of lexical signs) intended for use in automatic processing tasks. However, it remains relatively small, with 445 classes.

In general terms, sign spotting methods rely primarily on a signary when available, or on parallel data composed of videos and subtitles, where subtitles supervise the sign spotting process. One limitation of this second approach is that the correspondence between signs and words in subtitles is not straightforward. Some studies have already addressed this issue by incorporating word synonyms to enrich the possible links between signs and words in subtitles (Momeni et al., 2022).

In this work, we propose a spotting strategy that builds on named entities[1] in the subtitles to increase the number of signs spotted in the corresponding videos.

[1]real-world entities such as persons, locations, organizations, products, etc.

Our main contributions are threefold. (1) We automatically annotate a LSF dataset called Matignon-LSF with lexical signs using the pre-existing LSF-signary, and release these annotations to the community as the Matignon-signary. (2) We extend the dataset with named entity annotations. (3) We propose a new sign spotting approach that enriches existing lexical categories and uncovers novel sign classes, including rare signs.

The remainder of the paper is organized as follows. The section 2 reviews the related work in sign spotting and named entities in Sign Language Processing (SLP). Section 3 presents the preparation of the data used in the study. Section 4 describes the methodology and Section 5 reports the results.

2. Related Work

Sign spotting: Sign spotting aims to localize a specific sign in a continuous video. One common approach involves using a signary of isolated sign examples as queries. In this method, each sign in the signary, represented by a short video, serves as a reference to search for a similar sign within continuous video sequences. The search can be performed by computing similarity scores between visual representations of the query signs and the target video; the maximum similarity value is used to localize potential occurrences of the sign (Momeni et al., 2022). When multiple occurrences are available for a given sign, a classifier can be trained and used to detect signs directly (Jiang et al., 2021). Additional cues, such as mouthing, can be leveraged to improve spotting performance (Varol et al., 2022). When no signary is available, subtitles can serve as weak

Proceedings of the 12th Workshop on the Representation and Processing of Sign Languages, pages 203–211
16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

supervision by providing textual cues that indicate which video segments are likely to contain a given sign (Momeni et al., 2022; Lascar et al., 2024).

Named entities recognition: A named entity (NE) is a sequence of words or signs that refers to an identifiable entity in the real world, such as a person, a location or an organization. Named entity recognition (NER) is the process of detecting and labeling real-world entities in text. Several methods have been proposed in Natural Language Processing (NLP) (Keraghel et al., 2024). Recognizing named entities is crucial for enabling tasks like machine translation, question answering, and information retrieval. Although the field of sign language processing (SLP) has not been explored as extensively as NLP, several authors have nevertheless pointed out that named entities present significant challenges for translation (Camgöz et al., 2020), particularly with regard to localization or anonymization (Bleicken et al., 2016; Yin et al., 2021). Yin et al. (2021) pointed out that, in SLs, named entities may be conveyed either through fingerspelling, lexical sign, or by mouthing the name while pointing the spatialized referent with the index finger. The present research aims to determine how NE are signed in LSF and whether detecting NE in the subtitles provides useful cues for sign spotting.

3. Datasets and Data Preparation

3.1. Datasets

Two datasets are used in this work, Mediapi-RGB and Matignon-LSF, composed of LSF videos with the corresponding subtitles. They are respectively coming from native LSF (then translated into French subtitles), and interpreted LSF (from a spoken content). They are currently the largest corpora available for SLP in LSF.

The **Mediapi-RGB** (Ouakrim et al., 2024) corpus originates from Médiapi[2], a bilingual online media established in 2018. It contains news and reports in LSF produced by deaf journalists and subtitled in French, covering the period from 2018 to 2022. The corpus contains 86 hours of LSF video data, segmented into 50,084 video excerpts (average duration: five seconds), each aligned with its respective French subtitle (average length: 12.2 words).

Lexical units were extracted from the Mediapi-RGB corpus using the method described by Lascar et al. (2024). This process produced approximately 15k video instances of lexical signs distributed across 445 classes, covering various semantic categories such as locations, numbers, sports, and politics. Each video clip has a resolution of 444×444 pixels, a frame rate of 25 fps, and contains between 3 and 46 frames (average: 12.5 frames). In this work, we refer to this collection of lexical signs as the *Mediapi-signary*.

Matignon-LSF (Halbout et al., 2024) is an interpreted LSF dataset, containing 39 hours of video, and a vocabulary size of 10k subtitles across 18k sentences and 450k tokens. It features 15 LSF interpreters. It comes from the French government's Council of Ministers debriefings (from December 2020 to December 2023) and can be further extended[3]. Videos and subtitles are not aligned, there is a delay of a few seconds. For the sake of annotation, the sentence-like video segments are extracted following the subtitles sentences timestamps. This results in 18,173 videos. For our work, a 5-seconds margin is added at the beginning and at the end of each video to ensure good correspondence with the subtitle. This dataset is used in our study for signs spotting.

3.2. Data Preparation

We pre-processed the videos and subtitles, and identified the signers automatically to accommodate signer-dependent variations in signs.

Video preprocessing. All videos were encoded using a Video Swin Transformers (VST) model (Liu et al., 2022), initialized with publicly available Kinetics-400 pretrained weights (Kay et al., 2017) and subsequently fine-tuned on BOBSL (Albanie et al., 2021). A sliding window of 16 frames with a stride of 1 was used to encode the videos. Each sign of the Mediapi-signary is represented by a sequence of 16 frames, using zero-padding when the sign is shorter or truncation when it is longer, which occurs rarely. The 16-frames videos are then preprocessed the same way, and we obtain a single vector for each video.

Subtitles preprocessing. The Matignon-LSF's subtitles were normalized by lowercasing the text, converting cardinal numbers to digits, and expressing ordinal numbers in words.

Signer recognition. Unsupervised clustering was used to identify each signer in the Matignon-LSF dataset. We extracted the first frame from each video and embedded it using FaceNet512, an improved version of FaceNet (Schroff et al., 2015) ranked first in the DeepFace benchmark (Firmansyah et al., 2023). Embeddings were reduced to 100 dimensions via PCA and clustered with HDBSCAN (McInnes et al., 2017). The most relevant clusters were manually selected, and embeddings were classified using a Euclidean distance threshold of 1.3. When new signers appeared among unclassified videos, new classes were manually created. Iterating this process while changing the

[2] https://www.media-pi.fr/

[3] Debriefings of the Council of Ministers take place once a week.

frame representing each video led to the identification of 16 signers.

4. Sign Spotting Methods

Three methods have been used for automatically annotating the videos, illustrated by Fig. 1. Lexical signs were annotated using sign spotting from Mediapi-signary examples (Fig. 1 (a,b)), as explained in section 4.1. A second sign spotting technique (Fig. 1 (c)) consisted in pairing features from videos the subtitle of which contain a similar word (section 4.2). This technique was applied for NER, whose annotation is described in section 4.3. Section 4.4 proposes to use NE characteristics to better target the search (Fig. 1 (d)), allowing to collect rare signs.

4.1. Signary Spotting

The method is inspired from Momeni et al. (2022) and illustrated by Fig. 1 (a) and 1 (b). As Matignon-LSF does not contain annotation of mouthing cue, we focus on the subtitles as weak supervision for the spotting. From the Mediapi-signary, we select one class composed of a label, that is a French word, e.g. “sept” (*eng: seven*), associated to a collection of videos (e.g. the video A on Fig. 1 (a)). We use the label as the query to search within the subtitles. In Matignon-LSF, we select each video the subtitle of which contains the query word, called token *t* (e.g. “sept”). The hypothesis is that the corresponding lexical sign is present in video B. To find this occurrence, the features of A are compared to each feature of B using a sliding window. To that aim, the cosine similarity is computed between the VST features of A and B, which yields a similarity vector (represented by a heatmap in Fig.1).

A hysteresis thresholding is applied to segment the sign. The query sign A is considered to be found in B when a similarity score exceeds 0.5. Then, to extract the full sign, the segment is expanded to the right and left as long as the neighboring similarity scores exceed 0.45.

4.2. Pairing Videos to Annotate New Signs

The second annotation method, illustrated by Fig. 1 (c) comes from (Momeni et al., 2022; Lascar et al., 2024). It was used to create the Mediapi-signary. In that case, no query example is available, only the subtitles are used as weak supervision.

Cosine similarity is computed between videos containing a similar token *t*, and between a set of negative samples. The negative samples are used to unmatched similar frames that are not our target token such as pointing gestures that are common in SL contents. The final similarity is obtained by maximizing the contribution of positive pairs and minimizing that of negative pairs. To handle differences in signing style, videos were clustered by signer prior to similarity computation. This reduces variability in sign forms and leads to more reliable results. At least 9 postive samples (containing the token *t*) and 27 negative samples (not containing the token t) are needed to discover new signs. Following recommendations from (Momeni et al., 2022), we remove the 5s video margins (see section 3.1) to have more accurate negative samples. The same hysteresis thresholding as 4.1 is applied. The sign is considered a possible spot when the resulting segment lasts at least 3 frames. The Fig. 1 (c) illustrates the methods for the target sign PAYS.

4.3. Named Entities Annotation

Both previous methods rely on the token *t* as weak supervision for the spotting. However, not necessarily all words in spoken (and written) french are signed in LSF. Similarly, some signs have no direct counterparts in the subtitles. We have investigated the named entities by analyzing the annotations performed on the LSF 40 Brèves dataset[4](Challant and Filhol, 2022), and confirmed our findings through qualitative observations on the Matignon-LSF dataset. It appears that NE can be introduced by an hypernym sign (COUNTRY/LOCATION for a country, COMPANY for a company etc.) that are absent of the textual content. In order to confirm and quantify these signing patterns, we carried out the following experiment. We focused on geopolitical named entities (GPE), compagnies (ORG) and person (PERS) entities. GPE offer relevant examples in a constrained domain and allows for rapid and efficient manual annotation in the subtitles. On the other hand, compagnies and person entities requires automatic annotation due to the open vocabularies.

For the sake of clarity in the rest of this article, we introduce the following notations. We use a French word in small capital letters to designate the LSF signs, quotation marks for the words or excerpts from the subtitles, and the word *label* as an index to designate the name of the corresponding class in the Mediapi-signary (e.g. PAYS (*country*) refers to the sign, “pays” to the word, and pays$_{label}$ to the Mediapi-signary class).

The hypothesis made is that entities may be introduced by specific signs depending on the category forming a specific pattern composed of a minimum of two successive signs and denoted NE-PATTERN: the first sign introduces the NE category and is

[4]1 hour of journalist content fully annotated, including complex linguistic structures

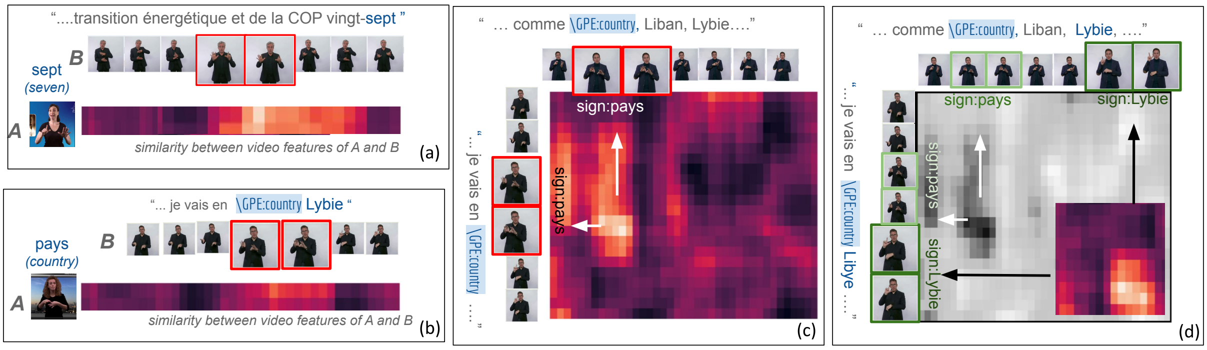

Figure 1: Illustration of the signary spotting (a,b) to find lexical signs associated to a word (a) or to a tag GPE:COUNTRY (b) in subtitles; (c) pairing to discover the signs corresponding to a word / tag present in the two subtitles; (d) targeted pairing where the sign PAYS (country) is detected and the sign corresponding to Lybia is searched locally with a higher precision.

more likely to be an hypernym, and the second refers to the entity. The realization of our example with the Lybia country would be written as follows: PAYS + LIBYE.

In this study, we have selected three GPE categories related to a country, a region and a department and two other categories: company and person. First of all, all the subtitles have been annotated manually with the following tags, written in capital letters: GPE:COUNTRY for a country, GPE:REG for an administrative region, GPE:DEP for an administrative department. For instance, the subtitle (translated in English) *"I am going to Libya."* will become *"I am going to GPE:COUNTRY Libya."*("Je vais en GPE:COUNTRY Lybie."). Concerning the companies (ORG) and persons (PER) annotation, to allow a rapid annotation, we use SpaCy[5] model, followed by a quick manual correction to remove major errors. We specifically labeled organizations (ORG) while restricting them as much as possible to companies[6].

To avoid biasing our study, we excluded the subtitles containing the words "pays", "région", "département", "entreprise" and "personne". For example, the subtitle *"I am leaving our **country** to go to GPE:COUNTRY Libya."* would be excluded ("Je pars de notre pays pour aller en GPE:COUNTRY Lybie."). This exclusion is motivated by the need to avoid duplicate spotting and to verify if the hypernym signs are present in the signing sequence while absent from the subtitles. Thus, 416 GPE:COUNTRY, 99 GPE:DEP, 58 GPE:REG, 2313 PER and 813 ORG were annotated [7] in the subtitles.

[5]https://spacy.io/, $fr_{c}ore_{n}ews_{s}mmodel$

[6]Examples of excluded entities include universities, NGOs, and government organizations, as they do not fall under the hypernym COMPANY

[7]Manual annotation was needed for GPE NE because SpaCy only provides general location NE, without distinguishing countries or departments.

Assuming that named entities can be explicitly introduced in LSF (e.g. PAYS for countries), the pairing method described in section 4.2 is leveraged using named entities tags to localize signs. Then, if these introducing signs belong to the Mediapi-signary, the signary-spotting of section 4.1 is used in a similar way as for lexical signs, but using a tag as query instead of a token. As illustrated by Fig. 1 (b), the lexical sign PAYS is used to find the sign associated to the GPE:COUNTRY tag. In the example, we look for PAYS whereas the word "pays" does not appear in the subtitle.

4.4. Targeted Pairing

The observation of the pattern sequence for the named entities inspired us a new method to extract new occurrences of signs, for example signs that are rare in the dataset and unknown in the Mediapi-signary. Indeed, the pairing method of section 4.2 preferably requires a minimum of 9 sign occurrences per class in the signary. Unfortunately LYBIE among others appears only three times in the subtitles. Thus, we propose to apply a *targeted pairing*, illustrated in Fig. 1 (d). The idea is to find the sign associated to the GPE tag (e.g., PAYS for GPE:COUNTRY) and to use it as an anchor. The sign of the entity is then searched by using pairing between two videos containing the sign to be found. Compared to the previous pairing described in section 4.2, the searching area is reduced in size, since the sign is searched right after the anchor. By reducing the area search, the search yields fewer ambiguities compared to the previous pairing, thus only 2 occurrences signed within the NE-PATTERN are enough instead of 9.

5. Results and Evaluation

After a description on the evaluation metrics in 5.1, section 5.2 provides the annotation results for tokens and NE. The section 5.3 focuses on the evaluation of two spottings methods: signary spotting and pairing videos. Finally, section 5.3 provides insight into the targeted pairing method and then focuses on the named entities pattern evaluation.

5.1. Metrics and Evaluation Procedure

First of all, duplicates caused by 5s video margins were automatically removed and visually verified. The number of reference annotations (ground truth **GT**) is obtained in 2 ways. When feasible (i.e. for GPE:REG and GPE:DEP for which the number of occurrences was limited), occurrences have been checked visually: $\mathbf{GT}_{vid}$. Otherwise, it is given by the number of words/tags in the subtitles: $\mathbf{GT}_{sub}$. Among the **N** signs detected, True Positive (**TP**) yields from visual verification by an LSF expert. Thus, the precision **P** is computed as **TP**/**N**. The recall $\mathbf{R}_{vid}$ (respectively $\mathbf{R}_{sub}$) is computed by **TP**/$\mathbf{GT}_{vid}$ (respectively **TP**/$\mathbf{GT}_{sub}$). The F1 score $\mathbf{F1}_{vid}$ is computed when available.

To go further, we also visually analyzed, for each GPE category, whether the signs spotted from the **TP** belong to its associated NE-PATTERN ($\mathbf{TP}_{pattern}$). When feasible (i.e. for GPE:REG and GPE:DEP), we generate a ground truth to indicate the presence of each pattern ($\mathbf{GT}_{pattern}$) in the video.

5.2. Quantitative Results

Token-based sign spotting. Using the sign spotting method described in section 4.1, *254* classes with **N**=*6,372* signs automatically extracted from Matignon-LSF. Those signs were evaluated by an expert, leading to **TP**=4,545, with a precision **P**=0.71 and $\mathbf{R}_{sub}$ =0.3.

This method does not account for ambiguities arising from the polysemy of certain signs. Consequently, relying solely on subtitles to detect signs is somewhat limited, which may partially explain the low $\mathbf{R}_{sub}$ scores. As an example, in French, the word “nord” (*north*) appears 24 times in the Matignon-LSF subtitles, with different meanings: referring to a French administrative department, a cardinal point, but also a part of the name of a French newspaper (“La voix du Nord”), 4, 11 and 9 times respectively. Using examples of nord$_{label}$ from Mediapi-signary, only 4 occurrences were retrieved, all related to the cardinal point. The newspaper and the department were signed differently, which illustrates the challenge posed by polysemy in sign detection.

Named-entities-based sign spotting. Since the subtitles have been tagged with NE categories, the pairing method of section 4.2 is used to retrieve the corresponding signs. Matching videos containing similar GPE tags results in finding signs corresponding to PAYS (9 occurrences), DÉPARTEMENT

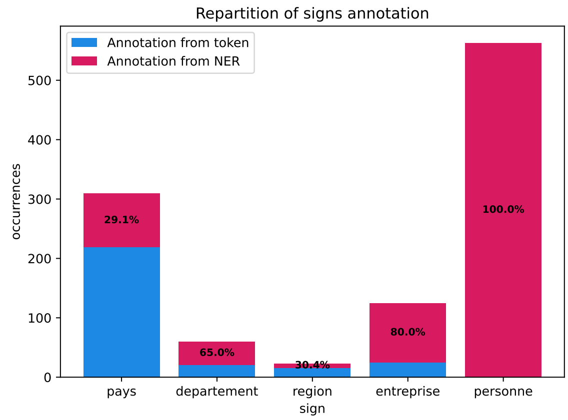

Figure 2: Number of signs after visual inspection (PAYS, RÉGION, DÉPARTMENT, ENTREPRISE and PERSONNE) detected using token (in blue) and NE (in red).

(5), RÉGION (2), ENTREPRISE (19) and PERSONNE (2). In a second stage, we use the classes pays$_{label}$, departement$_{label}$, region$_{label}$, entreprise$_{label}$ and personne$_{label}$ (present in Mediapi-signary) as queries for spotting (see section 4.1). The table 1 (last line) displays the results for all NE categories. The precision score (0.79) is high in contrast to the low recall $\mathbf{R}_{sub}$ score (0.21). This Table also reports the number of correct occurrences (**TP** = 768). By adding NE tags in the subtitles, the number of annotations is increased. The table also reports the results (indicated by an asterisk) of an alternative annotation approach from (Lascar et al., 2024), based on an MLP. More signs are correctly annotated for the ‘department’ and ‘region’ categories, but fewer signs are detected for the three other categories.

Fig. 2 displays, for each of the five categories, the proportion of signs spotted respectively using the token-based method (blue bar) and NE-based one (red bar). Respectively 29%, 30%, 65%, 80% and 100% more annotations are found for PAYS, RÉGION, DÉPARTEMENT, ENTREPRISE and PERSONNE. The signs RÉGION, DÉPARTEMENT were visually checked ($\mathbf{GT}_{vid}$) to compute F1 (0.78) and $\mathbf{R}_{vid}$ (0.7), which can be considered as solid.

5.3. Qualitative Evaluation

Discovering novel classes. Inspired by the qualitative analysis, we proposed a method in Section 4.4 to discover signs absent from the Mediapi-signary and thus increase the number of classes. The approach uses category signs (e.g., country, department) as anchors to perform targeted searches for subsequent signs. New classes must meet three criteria: absence from the Mediapi-signary, at least two NE-PATTERN annotations, and fewer than nine subtitle occurrences. We applied

sign	$\mathbf{GT}_{sub}$	$\mathbf{GT}_{vid}$	N	TP	P	$\mathbf{R}_{sub}$	$\mathbf{R}_{vid}$	$\mathbf{F1}_{vid}$	N^*	TP^*
PAYS	416	NA	115	90	0.78	0.22	NA	NA	13	13
RÉGION	58	10	8	7	0.88	0.12	0.70	0.78	66	56
DÉPARTMENT	99	54	45	39	0.87	0.40	0.72	0.79	189	158
PERSONNE	2313	NA	647	563	0.87	0.24	NA	NA	2	0
ENTREPRISE	813	NA	143	100	0.70	0.12	NA	NA	108	86
total	3699	NA	967	798	0.79	0.21	NA	NA	378	313

Table 1: Evaluation of the spotting of the GPE category signs. $\mathbf{GT}_{sub}$ is the number of tags in the subtitles, $\mathbf{GT}_{vid}$ is the number of GPE category signs in the video (manually annotated), **N** the total number of signs spotted (True Positive + False Positive) and **TP** the true positive. The scores **P** refer to the precision, $\mathbf{R}_{sub}$ to the recall from subtitles (GT_{sub}) and $\mathbf{R}_{vid}$ to the recall from the video ($\mathbf{GT}_{vid}$). The $\mathbf{F1}_{vid}$ is the F1 score with **P** and $\mathbf{R}_{vid}$. Results with * are obtained using an alternative MLP detector.

this method to two countries (Libya, Belarus) and one department (Landes). By computing cosine similarity and locating the nearest similarity blocks after the anchors, we retrieved several occurrences of these signs.

Country geopolitical entity. True positives of PAYS were manually analyzed to assess their conformity to the GPE:COUNTRY-PATTERN. Among 90 true positives, 61 matched the pattern, while the remaining 29 require further investigation. Based on 416 GPE:COUNTRY tags in the subtitles, only 14.67% followed this pattern. No full ground truth was established; instead, partial manual checks were conducted. Some countries are signed with specific signs, while others follow the GPE:COUNTRY-PATTERN by introducing the entity by the sign PAYS. Rarely mentioned countries (e.g., Lebanon, Iran, Belarus, Serbia) consistently follow the pattern, whereas frequently referenced countries, such as France's neighbors, influential states, or countries prominent in the news, tend not to. Several limitations were identified, including missed detections, list effects, fingerspelling, absence of signs, distinctive signs, and automatic filtering due to time margins. These factors may explain missing cases in the detected pattern. Overall, well-known countries appear less likely to be introduced by the sign PAYS and therefore less likely to conform to the GPE:COUNTRY-PATTERN. Future work should investigate the factors influencing the use of this pattern.

Region geopolitical entity. For GPE:REG, GPE:REG-PATTERNS were manually annotated in the corresponding videos, constituting the ground truth. Among the 7 TP occurrences, 5 of them conform to the GPE:REG-PATTERN, forming the $\mathbf{TP}_{pattern}$. Results are display in Table 2. The two remaining RÉGION are signed to add context in the signing sequences.

Department geopolitical entity Here also a GT is created for GPE:DEP-PATTERNS. Table 2 shows that nearly 31 DÉPARTEMENT out of the 99 of the GPE:DEP annotations are introduced by the hypernym sign DÉPARTEMENT. Otherwise, DÉPARTEMENT is used to provide additional context.

According to our observations, departments are signed in three ways: 1) use of its code[8] (70 occurrences among which 31 introduced by the sign DÉPARTEMENT); 2) fingerspelling (2 occurrences systematically introduced by DÉPARTEMENT); 3) a specific sign (12). The absence of the DÉPARTEMENT sign before the department's code is partly explained when several departments are listed, as shown by Table 3.

Company and person entity No complete manual observation was made for this two categories. **Company entity**: By inspecting references to SNCF (national railway company), we were able to retrieve the ORG-PATTERN. Out of the 12 mentions in the datasets, seven of them were preceded by the sign ENTREPRISE. Two of the mention in the subtitles are not signed (due to the context where the company has been already mentioned). Finally, three of the mentions in the subtitles are just signed with GARE (station) or fingerspelled, so without the scope of the ORG-PATTERN. Nevertheless, entities annotations remains partial, and the results could be improved through more thorough labeling. **Person entity**: This sign has not been spotted using only the subtitles. However, we were able to annotate more than 500 occurrences within the automatic named entities annotations. Due to the amount of annotation, we only analyze the PER-PATTERN for the entity *French President*: 215 out of the 725 mentions of the french president are preceded by PERSONNE. Further analysis could lead to understand how and when the pattern is used to mention someone.

6. Conclusion

This paper investigated three categories of named entities (locations, organizations and persons) and the specificity of their representation in French Sign Language (LSF), in particular the use of a "department" sign followed by a number for most French departments. This study provides several promis-

[8]From 01 to 95 for hexagonal France

NE-PATTERN	TP$_{pattern}$	GT$_{pattern}$	GT$_{sub}$
GPE:COUNTRY-PATTERN	61	NA	416
GPE:REG-PATTERN	5	7	5
GPE:DEP-PATTERN	31	39	99

Table 2: Number of retrieved patterns (**TP**$_{pattern}$) compared with the manually annotated ground truth (**GT**$_{pattern}$) for the NE-PATTERN

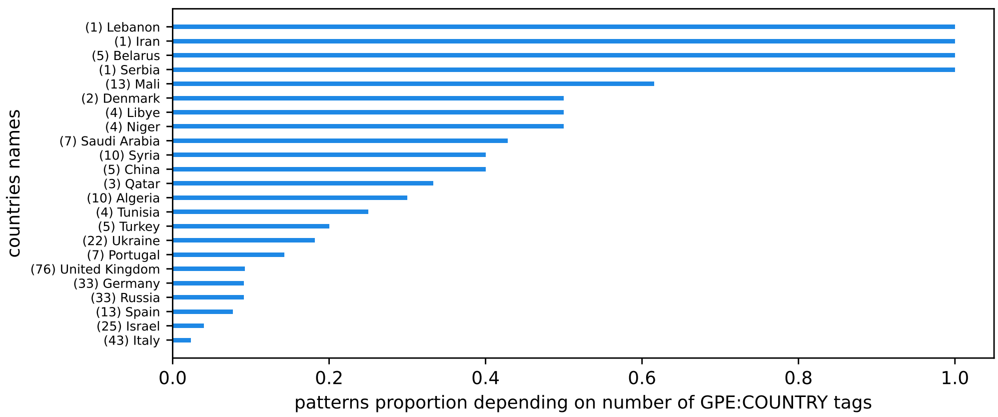

Figure 3: Proportion of GPE:COUNTRY-PATTERN: Number of GPE:COUNTRY-PATTERN instances over the total number of named entities GPE:COUNTRY in the subtitles.

Table 3: A subtitle with a list of departments from the V6VNZD5jLH8_segment_14 video excerpt: green ones conform to the GPE:DEP-PATTERN, blue ones are signed without being preceded by DÉPARTEMENT.

L'amélioration est continue, proche de la tendance nationale, dans l'Aube/GPE:DEP/NUM:10/GPE:DEP-PATTERN, dans le Rhône/GPE:DEP/NUM:69, en Île de France/GPE:REG, plus marquée dans la Nièvre/GPE:DEP/NUM:58, dans l'Eure/GPE:DEP/NUM:27, dans les Alpes Maritimes/GPE:DEP/NUM:06, et au contraire un peu plus lente en Seine Maritime/GPE:DEP/NUM:76, dans l'Aisne/GPE:DEP/NUM:02 et dans le Nord/GPE:DEP/NUM:59.

ing research directions: 1) enriching video annotations by enabling the detection of signs that are not present in subtitles (e.g., "department"). Our experiments showed an increase ranging from 30% to 100% depending on the category; 2) disambiguating certain translations from LSF into French (Ouakrim et al., 2024), where "department 17" can be correctly interpreted as "Charente-Maritime" rather than "17 departments"; 3) targeting the automatic retrieval of rare signs.

Finally, this work enabled a first step toward automatic annotation of the Matignon-LSF corpus, by exploiting both the lexicon present in the subtitles and our analysis of named entities.

7. Acknowledgements

This work was performed using HPC resources from GENCI-IDRIS (Grant 2024-AD011015441).

8. Ethical Consideration

Matignon-LSF is an interpreted sign language dataset. Recently, (Desai et al., 2024) pointed out the ethical problem of using interpreted dataset in the development of AI models since they differ from native SL and can lead to inaccuracies. Their use are nonetheless helping in the context of low-resourced data as those data are easy to collect, but the difference as to be acknowledged. Note also that none of the authors is deaf but some of them use LSF on a daily basis with Deaf colleagues, and the team's projects are carried out using an interdisciplinary approach (computer science and linguistics) and in collaboration with deaf organizations.

9. Bibliographical References

Julian Bleicken, Thomas Hanke, Uta Salden, and Sven Wagner. 2016. Using a language technology infrastructure for German in order to anonymize German Sign Language corpus data. In *International Conference on Language Resources and Evaluation (LREC 2016)*, pages 3303–3306, Portorož, Slovenia. European Language Resources Association (ELRA).

Necati Cihan Camgöz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020. Sign language transformers: Joint end-to-end sign language recognition and translation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Camille Challant and Michael Filhol. 2022. A first corpus of AZee discourse expressions. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1560–1565, Marseille, France. European Language Resources Association.

Mathieu De Coster, Dimitar Shterionov, Mieke Van Herreweghe, and Joni Dambre. 2023. Machine translation from signed to spoken languages: state of the art and challenges. *Univers. Access Inf. Soc.*, 23(3):1305–1331.

Aashaka Desai, Maartje De Meulder, Julie A. Hochgesang, Annemarie Kocab, and Alex X. Lu. 2024. Systemic biases in sign language AI research: A deaf-led call to reevaluate research agendas. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 366–377, Torino, Italy. ELRA Language Resources Association (ELRA) and the International Committee on Computational Linguistics (ICCL).

Diandra Fabre, Julie Lascar, Julie Halbout, Yanis Ouakrim, Annelies Braffort, Thomas Hueber, Michèle Gouiffès, and Denis Beautemps. 2025. Exploring Sign-level Strategies to Enhance Automatic Translation of French Sign Language. In *ACM Int. Conf. on Intelligent Virtual Agents*, Berlin, Germany.

Andrian Firmansyah, Tien Fabrianti Kusumasari, and Ekky Novriza Alam. 2023. Comparison of face recognition accuracy of arcface, facenet and facenet512 models on deepface framework. In *International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, pages 535–539.

Tao Jiang, Necati Cihan Camgöz, and Richard Bowden. 2021. Looking for the signs: Identifying isolated sign instances in continuous video footage. In *IEEE Face and Gesture*, page 1–8. IEEE Press.

Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, Mustafa Suleyman, and Andrew Zisserman. 2017. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*.

Imed Keraghel, Stanislas Morbieu, and Mohamed Nadif. 2024. Beyond words: A comparative analysis of LLM embeddings for effective clustering. In *Advances in Intelligent Data Analysis XXII - 22nd International Symposium on Intelligent Data Analysis, IDA 2024, Stockholm, Sweden, April 24-26, 2024, Proceedings, Part I*, Lecture Notes in Computer Science, pages 205–216. Springer.

Julie Lascar, Michèle Gouiffès, Annelies Braffort, and Claire Danet. 2024. Annotation of LSF subtitled videos without a pre-existing dictionary. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 100–108, Torino, Italy. ELRA Language Resources Association (ELRA) and the International Committee on Computational Linguistics (ICCL).

Ze Liu, Jia Ning, Yue Cao, Yixuan Wei, Zheng Zhang, Stephen Lin, and Han Hu. 2022. Video swin transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3202–3211.

Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205.

Liliane Momeni, Hannah Bull, K R Prajwal, Samuel Albanie, Gül Varol, and Andrew Zisserman. 2022. Automatic Dense Annotation of Large-Vocabulary Sign Language Videos. In *IEEE Computer Vision and Pattern Recognition*, volume 13695, pages 671–690. CVPR.

K R Prajwal, Hannah Bull, Liliane Momeni, Samuel Albanie, Gül Varol, and Andrew Zisserman. 2022. Weakly-supervised Fingerspelling Recognition in British Sign Language Videos. In *British Machine Vision Conference*, London, United Kingdom.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering . In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, Los Alamitos, CA, USA. IEEE Computer Society.

Gül Varol, Liliane Momeni, Samuel Albanie, Triantafyllos Afouras, and Andrew Zisserman. 2022.

Scaling Up Sign Spotting Through Sign Language Dictionaries. *International Journal of Computer Vision*, 130(6):1416–1439.

Gül Varol, Liliane Momeni, Samuel Albanie, Triantafyllos Afouras, and Andrew Zisserman. 2021. Read and attend: Temporal localisation in sign language videos. In *IEEE/CVF Computer Vision and Pattern Recognition*, pages 16852–16861.

Ryan Wong, Necati Cihan Camgöz, and Richard Bowden. 2023. Hierarchical i3d for sign spotting. In *Computer Vision – ECCV 2022 Workshops*, pages 243–255, Cham. Springer Nature Switzerland.

Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. 2021. Including signed languages in natural language processing. In *ACL-International Joint Conference on Natural Language Processing*, pages 7347–7360, Online. Association for Computational Linguistics.

10. Language Resource References

Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie Woll, Rob Cooper, Andrew McParland, and Andrew Zisserman. 2021. BOBSL: BBC-Oxford British Sign Language Dataset. *arXiv preprint arXiv:2111.03635*.

Julie Halbout, Diandra Fabre, Yanis Ouakrim, Julie Lascar, Annelies Braffort, Michèle Gouiffès, and Denis Beautemps. 2024. Matignon-LSF: a large corpus of interpreted French Sign Language. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 202–208, Torino, Italy. ELRA Language Resources Association (ELRA) and the International Committee on Computational Linguistics (ICCL).

Yanis Ouakrim, Hannah Bull, Michèle Gouiffès, Denis Beautemps, Thomas Hueber, and Annelies Braffort. 2024. Mediapi-RGB: Enabling Technological Breakthroughs in French Sign Language (LSF) Research through an Extensive Video-Text Corpus. In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: VISAPP*, volume 2, Rome, Italy.