Rosetta-LSF: an Aligned Corpus of French Sign Language and French for Text-to-Sign Translation

Élise Bertin-Lemée, Annelies Braffort, Camille Challant, Claire Danet, Boris Dauriac, Michael Filhol, Emmanuella Martinod, Jérémie Segouat

Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 4955–4962
Marseille, 20-25 June 2022

Rosetta-LSF: an Aligned Corpus of French Sign Language and French for Text-to-Sign Translation

Élise Bertin-Lemée[1], Annelies Braffort[2], Camille Challant[2], Claire Danet[2], Boris Dauriac[3], Michael Filhol[2], Emmanuella Martinod[2], Jérémie Segouat[4]
[1]SYSTRAN, 5 rue Feydeau, Paris, elise.bertinlemee@systrangroup.com
[2]LISN, Université Paris-Saclay, bât 507 rue du Belvedère, Orsay,
{annelies.braffort, camille.challant, claire.danet, michael.filhol, emmanuella.martinod}@lisn.upsaclay.fr
[3]R&D MocapLab, 70 rue du Landy, 93300 Aubervilliers, boris.dauriac@mocaplab.com
[4]STILS / CLLE, Université Jean Jaurès, 5 allée Antonio Machado, 31058 Toulouse, jeremie.segouat@univ-tlse2.fr

Abstract

This article presents a new French Sign Language (LSF) corpus called *Rosetta-LSF*. It was created to support future studies on the automatic translation of written French into LSF, rendered through the animation of a virtual signer. An overview of the field highlights the importance of a quality representation of LSF. In order to obtain quality animations understandable by signers, it must surpass the simple "gloss transcription" of the LSF lexical units to use in the discourse. To achieve this, we designed a corpus composed of four types of aligned data, and evaluated its usability. These are: news headlines in French, translations of these headlines into LSF in the form of videos showing animations of a virtual signer, gloss annotations of the "traditional" type—although including additional information on the context in which each gestural unit is performed as well as their potential for adaptation to another context—and AZee representations of the videos, i.e. formal expressions capturing the necessary and sufficient linguistic information. This article describes this data, exhibiting an example from the corpus. It is available online for public research.

Keywords: Sign Language, machine translation, example-based translation, synthesis

1. Introduction

This article presents a new French Sign Language (LSF) corpus called *Rosetta-LSF* (Dauriac, 2022). It has been designed with the main objective of allowing exploratory studies on the automatic translation of written French into LSF, output through virtual signer animation. A first study of this type is described in (Bertin-Lemée et al., 2022) for the translation part and in (Dauriac et al., 2022) for the animation part. The objective is to test an example-based approach, leveraging several types of aligned data.
The following section (section 2) specifies the framework of this study and explains a few choices made. Sections 3 and 4 respectively describe how the corpus was built and extended. Finally, section 5 describes in more detail an example taken from this corpus.

2. Text-to-Sign translation

The problem of automatically generating Sign Language (SL) from written language is one of translation, with text as the source format, in our case in French, and video or 3D animation as the target format, French Sign Language (LSF) for us. The meaning of the utterance must be preserved.
In line with the methodology of human translators and interpreters (with the exception of Sign Language interpreters) to preferentially be native in the target language to maximise the resulting fluency, the ultimate goal is clear comprehension in the target language.
In this section, we will look at the main challenges encountered in text-to-sign translation.

2.1. Constraints on data

In the case of spoken languages, machine translation (MT) was first developed in the 1950s using bilingual dictionaries and rule-based machine translation, with the idea of dealing with the grammar of source and target languages (word order, inflections, etc.) and the correspondence between them, in a controlled way. Although linguistically informed, this approach was later abandoned because developing and maintaining such a system of rules is extremely complex, especially when the linguistic structures are far apart, such as those of a spoken and a signed language. Moreover, a finite system is always insufficient to represent the generative power of language, especially in the case of multiple context-dependent translations, which are very frequent between French and LSF.

MT was largely transformed with available access to parallel corpora of examples. Statistical Machine Translation in the 1990s and 2000s used the frequencies of translation pairs containing source–target equivalent words or phrases in large human-translated corpora. This resulted in a breakthrough in translation quality. Today the dominant approach is Neural Machine Translation, which appeared in 2015 (Wu et al., 2016). Instead of an *a priori* model, it automatically determines, from the corpus, an intermediate representation in the form of numerical vectors that will allow a source text to be translated into a target text. It is therefore less directly explicable but the results are unquestionably better in quality compared to previous approaches.

However, both statistical and neural methods require very large volumes of parallel data (of the order of several million sentences). Whatever method is used for Sign Language Machine Translation (SLMT), the lack of large SL linguistic resources, and even more of bilingual corpora, hinders its development.

2.2. Intermediate representation

One of the major differences between SLMT and written MT is the difference in channel. As SL has no written form, SLMT requires a step to interpret or produce content in visual-gestural form. Thus, for text-to-SL MT, many approaches proceed in two steps: a first step maps the text content to an intermediate symbolic form representing the equivalent meaning in SL, and a second step uses this representation as the input of a synthesis system to animate a virtual signer.

This second step is not present in recent neural-based approaches that generate photo-realistic continuous sign videos from text inputs (Stoll et al., 2020). Like the other studies they are based on limited corpora, and to date, realism of the generated videos still needs to be improved. A major problem for us is that it does not offer anonymisation, unlike virtual signers whose appearance can also be adapted to suit the use case and audience. For this reason, avatar-based approaches seem more promising to date.

Among the few efforts using virtual signers, after a first generation of studies based mainly on the rule-based approach (Veale et al., 1998; Zhao et al., 2000; Marshall and Sáfár, 2004), some have investigated Example-Based Machine Translation (EBMT) (Morrissey and Way, 2005), sometimes combined with statistical approaches. For example, recently, De Martino et al. (2017) have developed a system that automatically translates Brazilian Portuguese text to Brazilian Sign Language (LIBRAS) by combining Statistical Machine Translation with EBMT to enable translations for unseen texts as well as translations of ambiguous terms dependent on the context and frequency of occurrence in previous translations. In general, these projects have not yet been completed and have not led to any follow-up, nor to consumer applications.

In addition, in the cases cited above, an intermediate representation is used consisting of gloss[1] sequences, each unit standing for a lexical unit generally restricted to manual activity. As such, they do not handle non-manual activity (or very few), spatial relations, or depicting structures. Yet, studies have shown that these structures can range from 20 to 70% of the units depending on discourse genre (dialog, storytelling, descriptions, etc.) (Sallandre et al., 2019). This results in incomplete and incomprehensible animations, which hinders their acceptability by the Deaf community. For this reason, it seems important to build bilingual resources that are linked by a richer intermediate representation than mere concatenations of glosses.

2.3. AZee as intermediate representation

We were able to show the advantage of *AZee*, a representation model for virtual signer animation (Hadjadj et al., 2018; McDonald and Filhol, 2021), particularly in terms of language coverage. A recent study has shown a minimum of 94% on a corpus of news items (Challant and Filhol, 2022). AZee is a formal approach for the representation of statements in LSF. It allows to define production rules that determine forms from semantic operations. By combining them, we build tree-structured expressions that generate signed statements while exposing their meaning.

For example, the expression below determines, thanks to the stacking of forms produced by each of the combined rules, the articulated forms for the translation of the statement "the president leaves quietly". It can be seen that the semantic combination of the operators involved in this expression is globally correctly interpreted with respect to the meaning of the complete source statement.

```
:info-about
    'topic
    :president
    'info
    :quietly
        'sig
        :leave
```

As this representation sufficiently determines the forms to be synthesized, it can be used as the output format of a translation system, which is more manipulable than attempting to render a video directly. To create a final rendering from produced AZee expressions, a synthesis step is required. The AZee output from the translation system is considered as the input to the synthesis system which produces the actual expected output, namely a sign language animation rendering. AZee is therefore a pivotal representation, and two types of alignment must be available.

Unlike a linear stream such as text or video where segments are delimited by a beginning and a length along a linear axis, an AZee expression is a hierarchical structure representing the hierarchy of the discourse constituents. Each node of the expression hierarchy therefore represents a portion of the utterance by itself, with the root node by definition covering the entire discourse. Thus, to align text with AZee expressions, the AZee "segment" side must take the form of a pointer to a single node of the expression, i.e. a source line number, instead of a start–duration specification used on the text side.

AZee–text alignments In order to translate written French into AZee expressions, we need a bank of alignments between each text segment and AZee expressions, each representing by definition a possible translation for the text segment. The "40 brèves" corpus (Filhol and Tannier, 2014) partly meets these needs because it provides text-video alignments, and the videos have been described with AZee: each French entry in the corpus is aligned with an AZee entry. But these data are few in number and the alignments were done at a coarse grain, at the discourse level. Finer-grained AZee alignment is necessary, for example to translate expressions of a few French words, or to synthesize isolated signs that would correspond to a word.

AZee–mocap alignments We intend to use the AZee expressions to animate the avatar from motion capture (mocap) data. Thus we also need an alignment between data recorded on real reference productions and the AZee expressions that represent them.

This led us to build a corpus offering the capacity to work on both fronts, on either side of the AZee pivot, which proves both a more comprehensive and flexible intermediate representation than glosses.

[1] A gloss is a text label, generally a single word, reflecting the meaning of the sign it stands for.

3. Constitution of the corpus

The corpus we present here was built in the framework of the French ROSETTA project, a French public/private project that studied accessibility solutions for audiovisual content. It included an exploratory study on automatic translation of subtitles in LSF displayed through signing avatars in a news broadcasting context. This section describes its design and recording conditions.

3.1. Choice of the LSF signer

We recorded the LSF productions of a Deaf person selected on the basis of her experience in producing LSF content for the media. She is a member of *Media'Pi!*, an independent news website, bilingual in French and LSF, and produces weekly news content for the website.

3.2. Elicitation material

The content of the corpus to be recorded and the documents to elicit this content were then established in such a way that the needs in translation and generation would be satisfied. We have defined four tasks of different nature:

Task 1 Translation of news titles
We have chosen this type of utterance as a case study for our project. News content exhibits well-formed and error-free language, deals with any topic, and is of varying size but never exceeding 30 words maximum. We have selected a list of near 194 news titles from the *France TV Info* French public information channel.

Task 2 Description of photos
This task elicited the production of depicting structures that are less common in news items. This was to enable a larger coverage of the language. We have selected 28 photos for this task.

Task 3 Reproduction of video clips
In order to have utterances with typical LSF linguistic structures, some involving specific non-manual aspects in particular, we selected video excerpts from priori LSF corpora, namely *40 brèves* (LIMSI, 2012), *DictaSign-LSF-V2* (LIMSI, 2020) and older *WebSourd* videos published online, and asked the signer to repeat them. For this task, 24 short videos were selected.

Task 4 Production of isolated signs
The most frequent lexical elements extracted from the titles of task 1 were isolated and completed with lexical elements from the evening news on *France 2*, a public TV channel, in order to generate variants of these titles. A list of near 1,200 words to be translated into LSF was prepared.

3.3. Technical set-up

The corpus was captured by motion capture thanks to two devices:

- a 37-camera optical motion capture system (Vicon) with retroreflective markers recording at 100 Hz;
- a head-mounted oculometer (MocapLab MLab 50-W) recording at 50 Hz.

Markers were placed on the whole signer's body, face and fingers allowing for a complete performance capture. A simultaneous infrared and RGB light characteristic signal was used to synchronise both systems.
An interpreter was facing the signer to discuss translation and two teleprompter screens were facing each of them. The elicitation material (instructions) was shown to both of them, translation was discussed on task 1 and 4, then the capture started. The teleprompter was not used during the shooting to avoid the signer's reading creating noise in the eye gaze direction, except for some elements in tasks 1 and 4 requiring finger-spelling (mainly proper names).

3.4. Nature of the corpus

After the motion capture, a 3D avatar with the same body proportions as the signer was created from the marker set. Its virtual "skin" is made of meshes. Its virtual "skeleton" is associated with the skin using a process called "rigging". The movement of the "bones" of this skeleton thus makes it possible to animate the skin of the avatar. Also, eye gaze was tracked and used to drive the eyes of the avatar. This type of rendering was chosen in order to obtain anonymised videos.
The avatar animations are then implemented into a 3D player developed on the Unity engine to produce a video for each acquisition, allowing a final rendering of the avatar (shades, lights, motion blur...).
For each element of the corpus except for task 4, a video of the avatar's front and right profiles from hips to head at 25 Hz was generated for annotation purposes as shown in fig. 2. This video was preferred to a real video during the shooting for several reasons: it is synchronised with the animation, its quality can be adjusted after the shooting, and it does not interfere with the motion capture shooting (i.e. lights on the studio and camera position in the acquisition volume that can result in marker occlusion).
In total, videos from task 1 lasted 38 min 30 s all together, 5 min 25 s for task 2, 3 min 10 s for task 3, and 2 h 14 min 22 s for task 4.

4. Extension of the Rosetta corpus

The next steps consisted in annotating the corpus to extend it, then creating the AZee–mocap and AZee–text alignments.

4.1. Corpus annotation

The annotation scheme was designed to allow the extraction of relevant elements for the generation process. In a conventional way, the manual units (including hand and arm activity) were segmented in the timeline. Then each unit was annotated on two tracks (right and left if needed, the signer being right-handed) with an attribute that allows to retrieve the elements to be reused from the mocap database and another one that describes the relation between the two hands if any.
While annotation is carried out in the classical way, by segmenting and annotating manual units, it is not limited to assigning a simple gloss to them. We indicate the different constraints to be applied on these units for any context of

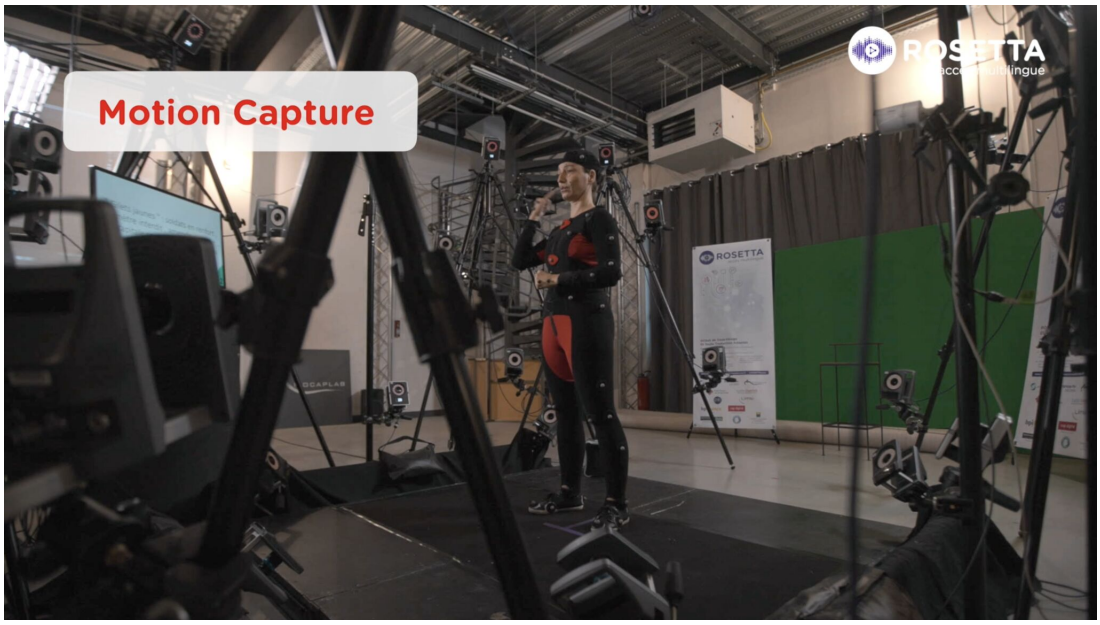

Figure 1: Motion capture setup

Figure 2: Avatar rendering

use. For each segment, three attributes have been specifically designed to help the generation process.

The first attribute is used to identify the articulatory constraints of the unit for the considered side, dominant or not. The objective is to indicate to the generation process the necessary and sufficient constraints, thus leaving the process free to modify the bendings of certain joints if necessary. This concerns the local constraints of all articulatory segments from the fingers to the shoulder. It is therefore more precise than what is usually called handshape. Note that no indication is given of the orientation and location of the sign, which is directly retrievable from the mocap data.

The second attribute describes the constraints on the performance of the considered articulator in relation to other parts of the body. The objective is to indicate the necessary and sufficient constraints to satisfy when modifications are applied to certain articulators (e.g. moving a hand, rotating the head, etc.).

The last attribute indicates the possibility or existence of constraints on the articulator with respect to the signing space. The objective is to indicate if the articulation depends on a spatial context (e.g. modification of orientation, location, amplitude), so that the generation can be adapted to the spatial context.

To date, all 194 titles in task 1 have been annotated.

4.2. Corpus AZeefication

The Rosetta corpus has also been used to create a bank of AZee discourse expressions—sometimes called AZee trees. This bank currently contains 194 AZee discourse expressions (see figure 3 for an example piece).

```
:info-about %F ont vendu leur vaisselle %t 1739-1967
   'topic
   :là
   'info
   :info-about
      'topic
      :all-of %F vaisselle %t 1767-1851
         'items
         list
            :assiette
            :assiette
      'info
      :multiplicity %F vendu %t 1855-1967
         'elt
         :vendre
```

Figure 3: Excerpt from the AZee discourse expression representing the LSF translation of the French news headline "*Samedi 30 et dimanche 31 mars, de grands chefs ont vendu leur vaisselle en Alsace, à Gerstheim.*" (Saturday 30th and Sunday 31st of March, top chefs sold their tableware in Alsace, in Gerstheim.)

Writing these expressions, a process we call "AZeefication", consists in identifying the AZee production rules and nesting them together, by observing the forms (involving manual and/or non-manual articulators) produced by the signer and interpreting their associated meaning. Moreover, the AZeefication of the corpus allowed us to participate in the stabilisation and adjustment of the model, by confronting it with a large data set.

The resulting set of AZee discourse expressions constitutes the intermediate data between the French text and the generation of animations. In other words the translation process will use AZee–text and AZee–mocap alignments, presented in the following sections.

4.3. AZee–text alignments

The generation of the LSF version by a virtual signer requires an alignment between the AZee discourse expressions and the French text. An AZee–text alignment is a correspondence between an AZee node (as above) and a segment in a textual news entry in written French. Each AZee discourse expression corresponding to a full text entry, a first set of AZee–text alignments can be formed, aligning each root node with the full AZee expression encoding it. But since each AZee expression is composed of sub-expressions and we may need various levels of granularity for the alignments, we also wish to align the sub-expressions (nested nodes) with smaller text segments when they can still be found. Like AZee–mocap alignments, AZee–text alignments are indicated in a comment, with pragma "%F", as shown in fig. 3.

Several principles have been followed in order to achieve consistent alignments that best reflect the structures present in LSF:

Uniqueness Within the same expression, the same segment in French cannot be aligned with different AZee sub-expressions (see example of what to avoid in Fig. 4.a).

Maximization If several alignments are possible as in Fig. 4.a, the solution with the largest sub-expression is preferred.

Objectivity Only a segment present in the news headline can be aligned with an AZee sub-expression, without escaping the translation instance hoping to extract generalities (see what to avoid in Fig. 4.b). Indeed, the purpose of EBMT is to generalise from specific examples, not to use example to encode generality.

When these three principles are respected, Fig. 4.c is the only one showing a correct alignment.

Once all of the AZee expressions in the corpus have been aligned following this method, an alignment file was created to collect them. Each alignment is encoded as follows:

- name of the text file in which the news title is found in French, e.g. RO1_X0007.Titre1;
- first and last characters of the aligned French segment, e.g. 10 4;
- file name of the AZee discourse expression, e.g. RO1_X0007.Titre1.az;
- line number of the aligned AZee expression or sub-expression, e.g. 7.

For example:

RO1_X0007.Titre1 10 4 RO1_X0007.Titre1.az 7

The AZee–text alignment file contains 1812 alignments.

4.4. AZee–mocap alignments

An AZee–mocap alignment is the correspondence between an AZee discourse expression and a video segment of the mocap corpus. This involves identifying the timed video segment covered by each AZee expression node. AZee-generated phenomena such as surrounding eye blinks or a hold of a manual configuration are taken into account in the alignment, which differs from the annotation performed in 4.1. To enable retrieval of a text segment from an aligned node, a "%t" pragma is appended on the AZee source line of that node, followed by the video frame numbers identifying the aligned segment (see fig. 2, top right, second line: 7713), as illustrated in fig. 3. We have only done the AZee–mocap alignments for the few video segments needed by the generation process, which explains why they are only included for a few examples in the downloadable corpus.

5. Example

To summarise, our corpus can be seen as a set of entries with aligned data of four different kinds. We return to the example of section 4, and present the four pieces of its entry:

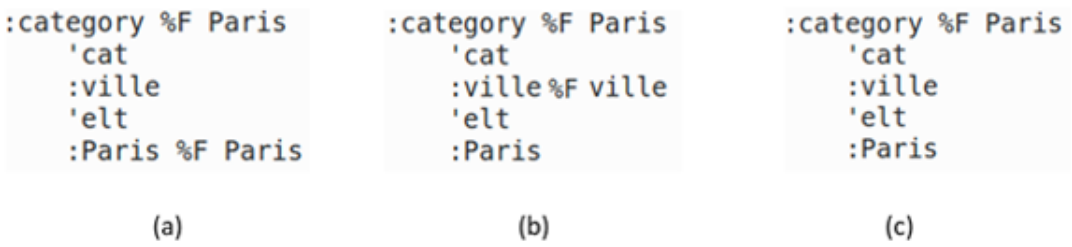

Figure 4: AZee expressions corresponding to the LSF translation of "Paris" in French with different AZee-text alignments

- the news title in written French, stored in RO1_X0069.Titre1;
- a 3D re-rendering of its LSF translation by a signer, after a full motion capture of the translation: RO1_X0069.Titre1.mp4;
- an annotation of the video in traditional glosses, as well as useful contextual information for the generation: RO1_X0069.Titre1.eaf;
- an AZee expression representing the signed discourse, containing all the text–AZee and some of the AZee-mocap alignments: RO1_X0069.Titre1.az.

Let us exemplify these four types of aligned data with an example of the corpus corresponding to the French expression: "*de grands chefs*" (top chefs). Its LSF translation is illustrated in fig. 5.

Table 1 shows the annotation. In addition to the usual attributes such as the start and end times, the indication of the dominant hand and the IdGlose, attributes specify important aspects for the generation:

- *personnes* and *chef_cuisinier* are bimanual symmetrical signs ("biequ" value for the `UnitT` attribute);
- for the three units, there are articulatory constraints on the handshapes up to the wrist (`Art` attribute labeled "wrist");
- there are internal dependencies with some articulators that are the other hand for *personnes*, the fingers for *pointage_main_circulaire* and the head for *chef_cuisinier* (`IntDep` attribute);
- we specify that the units *personnes* and *pointage_main_circulaire* are not performed in a canonical way in this extract (`ExtDep` attribut), and that this attribute is not applicable for the sign *chef_cuisinier*.

UnitT	Art	IntDep	ExtDep	IdGlose
biequ	wrist	otherh	notcanonical	personnes
mono	wrist	fingers	notcanonical	pointage_circ
biequ	wrist	head	notapplicable	chef_cuisinier

Table 1: Annotation of the extract

This supplementary information makes it possible to reuse the mocap extracts by adapting them if necessary to the context of the utterance to be generated.

Finally, the following AZee discourse expression completes the four-way alignment:

```
:category %F de grands chefs %t 1587-1727
    'cat
    :side-info
        'focus
        :multiplicity
            'elt
            :une personne
        'info
        :pointage zone %I plan horizontal
    'elt
    :chef cuisinier
```

The `category` rule carries the meaning: "*elt*, to be taken as an instance of *cat*". The top-level expression here therefore means *chef cuisinier*, i.e. the third unit "top chef", to be understood as an instance of the *cat*, which contains the first and second units, namely *une personne* and *pointage zone*, together meaning a group of people located in the signing space.

These units are connected through the `side-info` rule, which carries the meaning: "*focus*, with non-essential information *info* about it".

The `multiplicity` rule is applied to *une personne* ("a person"). It adds a notion of plural to the unit, thus its application here denotes a group of people. In other words, the `multiplied` *une personne* creates a group of people.

Then, the *pointage zone* shows where these individuals are represented in the signing space (`side-info` rule).

Finally, the *chef cuisinier* belongs to the larger category of *une personne* (`category` rule).

These AZee operations generate all the form specifications that enable the generation of animations of the overall utterance, whether manual or non-manual, as well as the appropriate temporal structuring.

This annotated LSF piece from the corpus could be then used to generate a new sentence such as: "*De grands chefs ont vendu leur vaisselle pour les plus modestes dans la banlieue de Gerstheim.*" (Top chefs sold their tableware for households in the lowest income group in the suburbs of Gerstheim.)

Figure 5: Snapshots from the video showing the three gestural units annotated with glosses implied in the LSF extract aligned with French text "*de grands chefs*" (top chefs), if we were to limit ourselves to the annotation of manual units

6. Conclusion

Aimed at experiments on automatic French-to-LSF translation of news content, the corpus we built in the Rosetta project consists of 194 news headlines and 1200 isolated words translated from French, as well as material focused on LSF-specific constructs, with 22 picture descriptions and 22 video reproductions. They were all recorded with motion capture and rendered by a virtual human character. Reflecting the visuo-gestual modality of sign languages to convey meaning, we provide for the news headlines part not only text and videos, but also rich gesture annotation to help a more fluent use of the recorded extracts in new generated contexts, as well as sentence and phrase-level semantic and structural annotations (AZee discourse expressions) based on the hierarchical AZee representation. These two types of annotations allow for a much better representation of the SL-specific phenomena, not conveyed when only glosses are used as an intermediate representation: multi-linearity, use of space and iconicity.

The Rosetta-LSF corpus is registered on the ISLRN website[2] and is available online for public research on the Ortolang website[3].

It was actually used in the framework of a project of automatic translation from written French into LSF, which led to the creation of a prototype. The description of this prototype and the corresponding architecture is outside the scope of this article, but the interested reader will find a proof-of-concept video on the project website[4], and we are working to publish papers describing those processes.

While it has fulfilled its role of enabling an exploratory study on example-based machine translation, the corpus remains too small for translation with a large span of French and LSF. It could of course be extended following the same protocols as those presented here, which requires time and human resources to carry out the annotations and AZeefications. One way that could be explored is to study how these annotations could be automated or assisted, and most importantly temporally synchronised in order to allow faster and cheaper production. The progress made in the field of automatic analysis and recognition in Sign Language videos could be leveraged.

[2]http://islrn.org/

[3]https://www.ortolang.fr/market/corpora/rosetta-lsf

[4]https://rosettaccess.fr/index.php/rosettas-final-demonstrator/

7. Acknowledgements

This work has been funded by the Bpifrance investment project "Grands défis du numérique", as part of the ROSETTA project (RObot for Subtitling and intElligent adapTed TranslAtion).

We thank Noémie Churlet, Raphaël Bouton and Media'Pi! for their commitment to this project, which would not have had the same validity and impact without them.

8. Bibliographical References

Bertin-Lemée, E., Braffort, A., Challant, C., Danet, C., and Filhol, M. (2022). Example-Based Machine Translation from Text to a Hierarchical Representation of Sign Language. *arXiv*.

Challant, C. and Filhol, M. (2022). A First Corpus of AZee Discourse Expressions. In *Language Resources and Evaluation Conference (LREC), Representation and Processing of Sign Languages, Marseille, France*.

Dauriac, B., Braffort, A., and Bertin-Lemée, E. (2022). Example-based Multilinear Sign Language Generation from a Hierarchical Representation. In *Sign Language Translation and Avatar Technology Workshop: The Junction of the Visual and the Textual*, Marseille, France.

De Martino, J. M., Silva, I. R., Bolognini, C. Z., Costa, P. D. P., Kumada, K. M. O., Coradine, L. C., da Silva Brito, P. H., do Amaral, W. M., Benetti, Â. B., Poeta, E. T., et al. (2017). Signing Avatars: Making Education More Inclusive. *Universal access in the information society*, 16(3):793–808.

Filhol, M. and Tannier, X. (2014). Construction of a French-LSF corpus. In *Workshop on Building and Using Comparable Corpora*, Reykjavík, Iceland, May.

Hadjadj, M., Filhol, M., and Braffort, A. (2018). Modeling French Sign Language: a Proposal for a Semantically Compositional System. In *International Conference on Language Resources and Evaluation*.

Marshall, I. and Sáfár, E. (2004). Sign Language Generation in an ALE HPSG. In *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar (HPSG-2004)*, pages 189–201, August.

McDonald, J. and Filhol, M. (2021). Natural Synthesis of Productive Forms from Structured Descriptions of Sign Language. *Machine Translation*, 35(3):363–386.

Morrissey, S. and Way, A. (2005). An Example-Based Approach to Translating Sign Language. In *Second Workshop on Example-based Machine Translation*.

Sallandre, M.-A., Balvet, A., Besnard, G., and Garcia, B. (2019). Étude exploratoire de la fréquence des catégories linguistiques dans quatre genres discursifs en LSF. *LIDL : Linguistique de Didactique des Langues*, 60: Langues des signes et genres discursifs.

Stoll, S., Camgoz, N. C., Hadfield, S., and Bowden, R. (2020). Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks. *International Journal of Computer Vision*, 128(4):891–908.

Veale, T., Conway, A., and Collins, B. (1998). The Challenges of Cross-modal Translation: English-to-Sign-Language Translation in the Zardoz System. *Machine Translation*, 13(1):81–106.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv preprint arXiv:1609.08144*.

Zhao, L., Kipper, K., Schuler, W., Vogler, C., Badler, N., and Palmer, M. (2000). A Machine Translation System from English to American Sign Language. In *Conference of the Association for Machine Translation in the Americas*, pages 54–67. Springer.

9. Language Resource References

Dauriac, B. et al. (2022). *ROSETTA-LSF corpus*. distributed via ORTOLANG: `https://hdl.handle.net/11403/rosetta-lsf/v1`, v1.

LIMSI. (2012). *40 brèves*. ISLRN 988-557-796-786-3.

LIMSI. (2020). *Dicta-Sign-LSF*. v2, ISLRN 442-418-132-318-7.