

Sign Language corpus analysis: Synchronisation of linguistic annotation and numerical data

Jérémie Segouat, Annelies Braffort and Emilie Martin

LIMSI/CNRS, Orsay, France
Campus d'Orsay, Bat. 508, BP133, F-91403 Orsay cedex
annelies.braffort@limsi.fr

Abstract

This paper presents a study on synchronization of linguistic annotation and numerical data on a video corpus of French Sign Language. We detail the methodology and sketches out the potential observations that can be provided by such a kind of mixed annotation. The corpus is composed of three views: close-up, frontal and top. Some image processing has been performed on each video in order to provide global information on the movement of the signers. That consists of the size and position of a bounding box surrounding the signer. Linguists have studied this corpus and have provided annotations on iconic structures, such as “personal transfers” (role shifts). We used an annotation software, ANVIL, to synchronize linguistic annotation and numerical data. This new approach of annotation seems promising for automatic detection of linguistic phenomena, such as classification of the signs according to their size in the signing space, and detection of some iconic structures. Our first results must be consolidated and extended on the whole corpus. The next step will consist of designing automatic processes in order to assist SL annotation.

1. Introduction

French Sign Language (FSL) is the visuo-gestural language used by the French deaf community. Research on the FSL, as for all Sign Languages (SL), requires building and analyzing video corpora.

Two multimedia annotation pieces of software are dedicated to SL corpora analysis: ILex (Hanke, 2002) is a tool for SL lexicography and corpus analysis allowing direct access to a lexicon stored in a database. SignStream (Neidle, 2002) allows multiple utterances to be open at the same time, permitting side-by-side comparison of data.

Two other multimedia annotation software are dedicated to video corpora analysis, but not especially for sign languages. Elan (Wittenburg, 2002) is not dedicated to SL, but the associated metadata tool provides a SL profile. Anvil (Kipp, 2001) is not dedicated to SL, but allows several speech tiers to be activated.

One of our aims is to participate in the development of such tools, which should include for example image processing tools, numerical analysis and recognition tools (Braffort, 2004). That includes studying which kind of automatic analysis applied to the video would enable us to extract information on the structure of utterances in FSL.

Thus, to identify the relevant treatments, we have started a study that consists in correlating linguistic annotation with numerical data, provided by a pre-process of the video. The idea is to use the speech tiers to visualize numerical data.

This paper describes the methodology and sketches out the potential observations that can be provided by this kind of mixed annotation.

2. Methodology

Several national multi-disciplinary projects dedicated to FSL have been initiated in France since 2000, where both linguists and computer scientists were involved.

During one of these projects, named LS-COLIN¹, a video database of FSL was built, with the double aim to

provide data for linguists who want to highlight the iconicity of the FSL, and to provide good quality videos for automatic analysis (Cuxac 2001).

This corpus is composed of several kinds of discourse on different topics: narrative (two different stories), explicative (cooking), argumentative (on important events), explicative meta-linguistic (on linguistic courses). Thirteen signers were recorded on each topic, and one of the story was performed twice by each signer, in order to study both inter and intra variations (Cuxac 2001).

This corpus was recorded with three cameras, providing three views: close-up, frontal and top (Figure 1). It was recorded at the INJS² Institute in Paris, with blue background, dark clothes, and high quality lighting, in order to allow researchers to perform image processing on the videos (Braffort, 2001), (Mercier, 2005).



Figure 1: The three views of LS-COLIN corpus

With such a kind of corpus, linguists and computer scientists can study the same video together, with the aim to perform complementary analysis.

¹ <http://www.irit.fr/LS-COLIN>

² INJS: Institut National des Jeunes Sourds

2.1. Linguistic annotation

The purpose of one of the linguistic analyses performed on LS-COLIN corpus, was to refine the categorization of the discourse units established by Cuxac (Cuxac, 2000). More precisely, the goal was to study in details the linguistic processes carried out when the aim of the signer is “to show what he is saying”. That is what Cuxac named “illustrative aim”. He has distinguished three kinds of linguistic structures that he called *transfers*: The *size and shape transfer* (TTF), which is used to describe the shape of a person, an object or a place (Figure 2a), the *situational transfer* (TS), which is used to show the displacement of a person or an object relatively to a stable locative reference (Figure 2b), and the *personal transfer* (TP), where the signer “becomes” one of the person or object of the discourse (Figure 2c). TS and TP can be combined in double transfers (DT), such as in Figure 2d. Sometimes, some parts of lexical signs can appear in DT (Figure 2e)

Sallandre carried out this work during her PhD (Sallandre, 2003). She enriched the classification by adding sub-categories in TP and DT. Therefore, she has annotated an important part of the LS-COLIN corpus, with the description of the iconic structures encountered. We have used her annotation in this study.



Figure 2: Examples of transfers

- a: A TTF, “spread pastry” in a recipe,
- b: A TS, “a bird on a fence” in the horse story (HS)
- c: A TP, “a horse galloping” in the HS
- d: A DT, “a ruminating cow” in the HS
- e: A DT with lexicon sign, “the cow (proform with the dominated hand) is waiting (dominant hand)” in the HS

2.2. Numerical annotation

An analysis of the video by image processing was also carried out on the same corpus by Cassel during his PhD (Cassel, 2005). He applied human detection and tracking

in a video, designed in the context of acrobatic movements. This process provides four temporal data: Position (X, Y) and size (L, H) of the bounding box which surrounds the signer, as shown in Figure 3. The software provides a set of curves for each video file, that we can use for analysis (Figure 4).

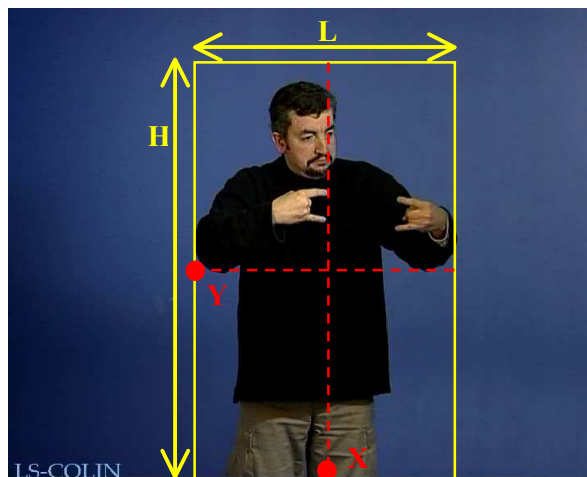


Figure 3: The bounding box and the 4 correlated values.

For example, in the frontal view, the bounding box tracks the signer’s body, head and arms: The width represents the distance between the hands or the shoulders, depending on the position of the hands, and the height represents the distance between the legs and the head, or the hands if they are above the head.

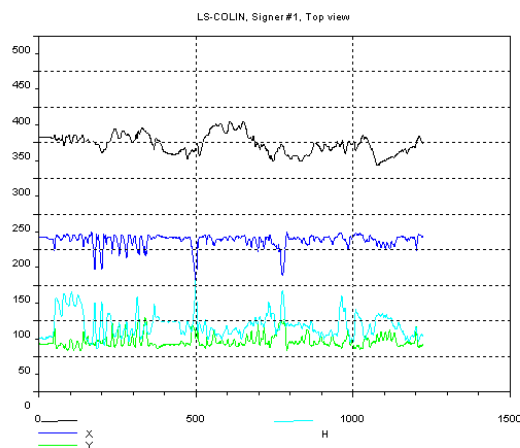


Figure 4: Example of a set of curves for the top view

The study has consisted in integrating these two types of annotation into an annotation software and to analyze the possible correlations between linguistic phenomena and numerical data. For the moment, only one FSL story, signed by two different persons, was annotated (that is two videos). The correlations observed were listed manually. The annotation tool used was ANVIL (Kipp, 2001) because at this moment this is the only tool that enables us to import the data resulting from the image processing, to display them as curves and to synchronize them with the linguistic annotation (Figure 5).

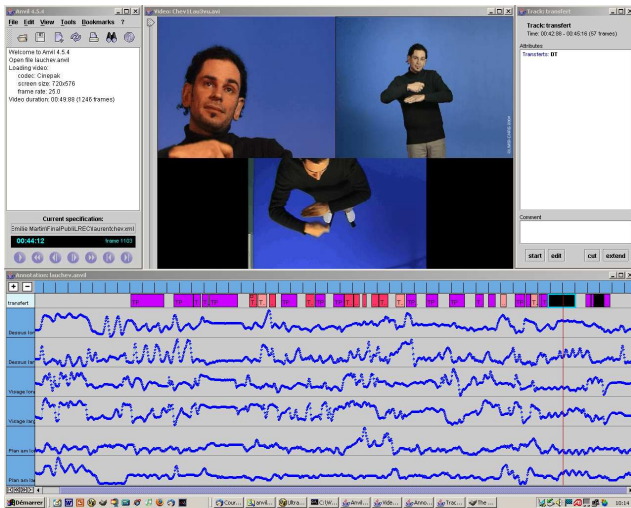


Figure 5: Example of annotation with linguistic and numerical data.

3. First observations

Our observations are partly qualitative, and partly quantitative.

A first observation is that the curves representing the width and the height of the bounding box can be used to measure the occupation of the signing space, by looking at the variations in the three views. This global information allows us to classify the signs according to the amplitude (in width and length) in the different views.

For instance, the H variable in the front view can be used to detect the signs for which the hands are above the head: The peaks in the read circle in the Figure 6 correspond to the sign shown in the Figure 7.

For the L variable in the same view, the low values (under the mean value) in the green circle in Figure 6 correspond to the moments where the elbows are oriented toward the center of the body (Figure 8), and the peeks above the mean value (blue circle in Figure 7) correspond to bimanual signs with hands outside the body (Figure 9) or monomaneal signs with the hand far from the body (Figure 10).

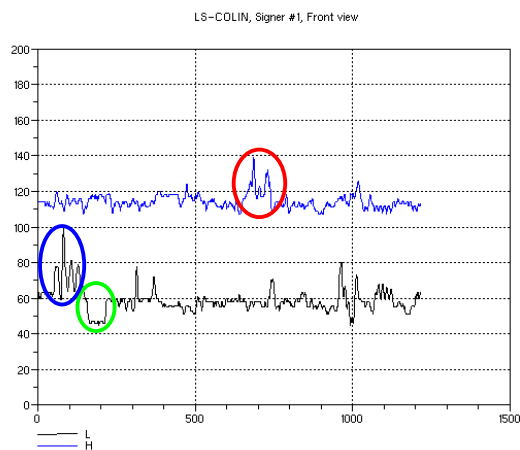


Figure 6: L and H curves for the frontal view.

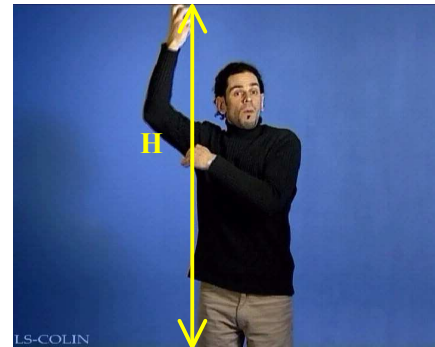


Figure 7: The horse is falling down (TS)

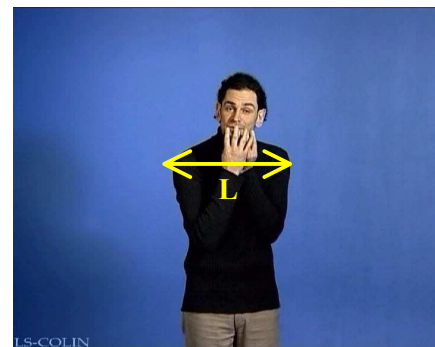


Figure 8: [FIELD]_{LSF}

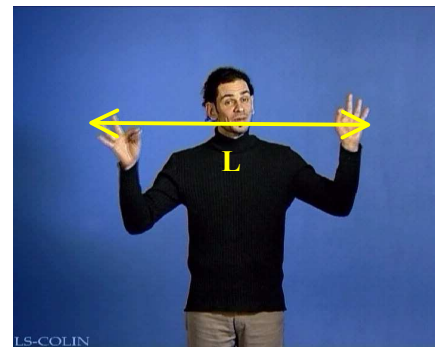


Figure 9: [STORY]_{LSF}

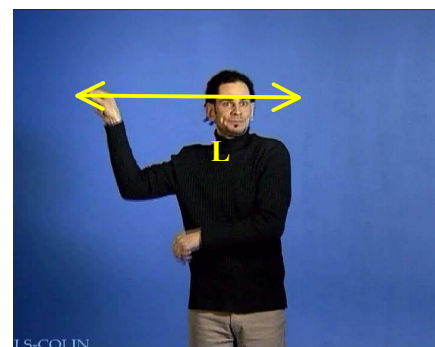


Figure 10: The bird is taking a first-aid kit (TS)

A second observation concerns the structures of iconicity annotated by the linguists. Some of the numerical data, in particular in the top and close-up views, seems relevant to detect certain types of transfers.

Thus, the first time a character is introduced in the story, a personal transfer (TP) is often performed during a significant period. These characters are located in the signing space on the left- or right-hand side of the signer. For example, in the cow story, the cow is located on the right (Figure 1) or on the left (Figure 11) in the signing space, depending on the signer.



Figure 11: TP 'cow' with the signer #2.

The displacement of the signer's body is perceptible on the X curve of the top view. In Figure 12, the low value of X in the red box corresponds to the signer's body shifts on the right, while the other curves keep a stable value.

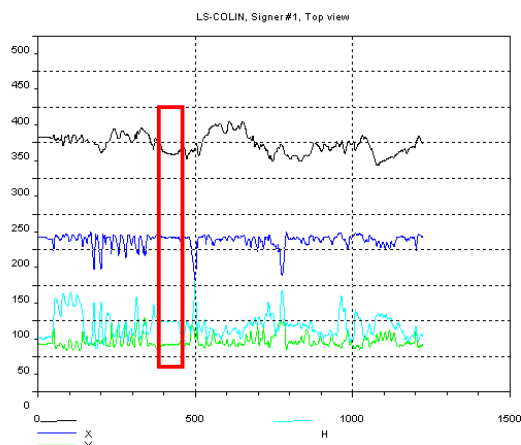


Figure 12: Detection of a right TP on the top view

The curves of the face view have also stable values during the same period (Figure 13).

4. Conclusion

All the results we present here must be refined and confirmed by extending the study on the whole corpus. In a second step, data that is more local, for example on the face and hands of the signer, should be added to the global data we used in this first study. Detection and classification would then be more accurate.

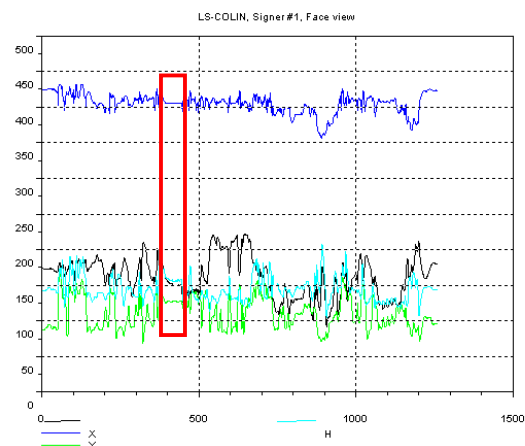


Figure 13: Detection of a TP on the close-up view

For this first attempt in correlating linguistic and numerical annotations, even if the results are only outlined, they are promising and should help to develop automatic tools for annotation and classification in the field of SL video corpora based on visual cues. This study consolidates us in our multidisciplinary approach of annotation including image processing and automatic analysis. Our first results must be consolidated and extended on the whole corpus. When the relevant treatments will be sufficiently defined and validated, they will be integrated in the annotation software dedicated to video SL corpora (Braffort, 2004).

5. References

Cuxac C. et al. (2001) Projet LS-COLIN. Quel outil de notation pour quelle analyse de la LS ? RLSF'01.

Braffort A., Choisier A., Collet C. et al. (2004) Toward an annotation software for video of Sign Language, including image processing tools and signing space modelling. LREC 2004.

Cassel R., Collet C., Gherbi R. "Real-Time Acrobatic Gesture Analysis", 6th International Workshop on Gesture in Human-Computer Interaction and Simulation, 2005, Vannes, France.

Cuxac C. (2000) La Langue des Signes Française (LSF) – Les voies de l'iconicité. In: "Faits de Langues" 15-16, Ophrys.

Hanke, T. (2002) iLex - A tool for sign language lexicography and corpus analysis. LREC 2002.

Kipp M. (2001) Anvil - A Generic Annotation Tool for Multimodal Dialogue. In Eurospeech'01, 1367-1370.

Neidle, C. (2002), SignStream™: A Database Tool for Research on Visual-Gestural Language. In B. Bergman, P. Boyes-Braem, T. Hanke, and E. Pizzuto, eds., "Sign Transcription and Database Storage of Sign Information, a special issue of Sign Language and Linguistics" 4:1/2, 203-214.

Sallandre M.-A. (2003) "Les unités du discours en Langue des Signes Française. Tentative de catégorisation dans le cadre d'une grammaire de l'iconicité." PhD thesis, Paris 8 university.

Wittenburg P., Brugman H., Levinso, St., Kita S. (2002). Multimodal Annotations in Gesture and Sign Language Studies. LREC 2002.