Sign Language corpora for analysis, processing and evaluation

A. Braffort, L. Bolot, E. Chételat-Pelé, A. Choisier, M. Delorme,
M. Filhol, J. Segouat, C. Verrecchia, F. Badin, N. Devos

LIMSI/CNRS

Orsay, France

E-mail: annelies.braffort@limsi.fr

Abstract

Sign Languages (SLs) are the visuo-gestural languages practised by the deaf communities. Research on SLs requires to build, to analyse and to use corpora. The aim of this paper is to present various kinds of new uses of SL corpora. The way data are used take advantage of the new capabilities of annotation software for visualisation, numerical annotation, and processing. The nature of the data can be video-based or motion capture-based. The aims of the studies include language analysis, animation processing, and evaluation. We describe here some LIMSI's studies, and some studies from other laboratories as examples.

1. Introduction

Sign Languages (SLs) are the visuo-gestural languages practised used by deaf communities. The number of body features involved in SL communication allows for a lot of information at once: SLs do not only use hands, but also shoulders, eye gaze, facial expression, head movements (figure 1); linguistic studies of SL show a heavy and consistent use of the "signing space", i.e. the portion of space in which the signs are performed (Liddell 2003); iconicity is also a relevant feature of both its lexicon and its grammar (Cuxac 2000).

Research on SLs requires building and analysing corpora. These corpora are mainly composed of video, which can be constituted of several shots, in order to capture the 3D nature of the language (figure 1). The classical use of these video corpora is to study SL functioning. In Sign Language Processing domain, the aim is to elaborate formal representations of linguistic phenomena, and implement them in computer applications, which can be automatic recognition (Drew et al. 2007), generation (Elliott 2007), and translation (Morissey 2008). The nature of the data (3D and multilinear) implies to look for new ways of collecting knowledge, by using new methods, numerical data or specific processing.

The aim of this paper is to present various studies that explore new kinds of data or new uses that can be made of present SL corpora with the aim to provide additional knowledge than the sole video. These studies are a selection of our studies at LIMSI laboratory and studies at other laboratories, which shows the variety of the attempts. They are divided in three parts: video corpora for language analysis, video and 3D corpora for animation processing, and video corpora for evaluation.

2. Corpus for language analysis

In this section, we present three examples of studies using video or 3D SL corpora for analysis. These studies are based on original methods with the aim to capture more phenomena or to compensate the lack of 3D data. We first present two examples of studies conducted at LIMSI laboratory, using video LSF (French Sign Language) corpora for analysis on coarticulation, and on non-manual gestures. The last example shows the use of numerical data provided by a dataglove into annotation software.

2.1. Simultaneous visualisation of variations

The first study concerns the coarticulation phenomenon (Segouat 2009). Coarticulation is the set of modifications that occur between and within signs when performed in utterances, as opposed to isolated realisations. To analyse these modifications, we have recorded several videos of isolated signs, and utterances that involve those signs. We have annotated both signs and utterances and compared these annotations to reveal the modifications due to coarticulation.

The complete utterance and the isolated signs are composed, both in the video display (figure 1) and the annotation board (figure 2) of the software, in order to process analysis. We use alignment and statistical methods to extract correlations, in order to design a formal model of coarticulation.

In figure 3, we can notice that there are slightly different facial expression, gaze, and hands location, depending on whether the sign is performed isolated (right side of the figure 3) or within an utterance (left side).
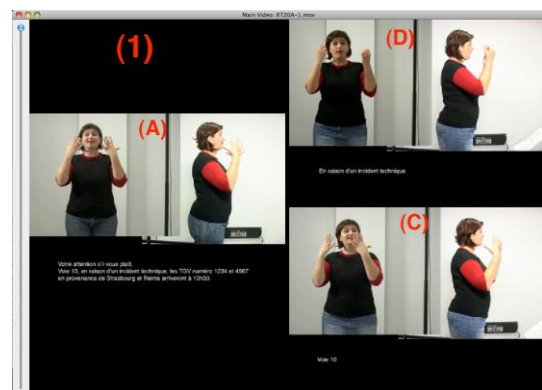

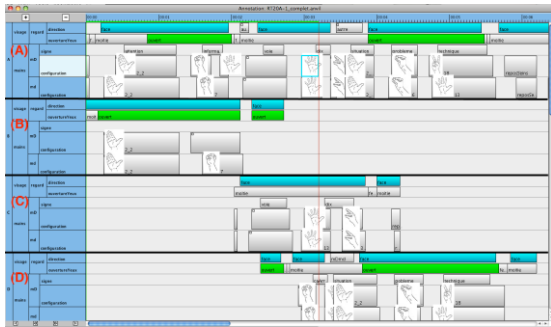Figure 1: Simultaneous visualisation in ANVIL. WebSourd corpus.

Figure 2: Simultaneous annotation in ANVIL.

2.2. Direct annotation on the video

The purpose of the second study is to model the non manual gestures (NMGs) involved in LSF. We have undertaken a very fine annotation of four videos extracted from the LS-Colin corpus (Braffort et al. 2001). This corpus is characterized by a high level of image quality and provides a close-up view which is particularly precious for this study. Our annotation methodology allows for a fine description of movements and takes the temporal structure of NMGs into account, in particular eyebrow movements and eye blink (Chételat-Pelé et al. 2008). We use statistical methods to define a list of recurrent structures for each NMG and the linguistic context where they occur. The resulting formalisation will be used for automatic generation.

Figure 3 illustrates the way eyebrow movement is described: 18 points are manually positioned on each frame of the video (with a rate of 25 frames/second). The variation of the eyebrow posture is computed related to the eye corner locations. This allows the eyebrow movements to be numerically described.

Figure 3: Annotation of eyebrow movements in ANVIL. LS-COLIN corpus.

Figure 4 shows the set of symbols used to annotate the eyelid movements. For these elements, we only describe the movement phases (raise, lower, stable). In this example, the green block represents eye blinks, the yellow one to wrinkled eyes, and the blue ones raised eyebrows and eyes wide open.
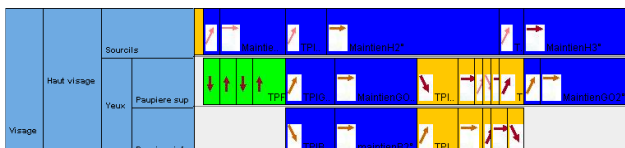

Figure 4: Annotation of eyelid movements in ANVIL.

2.3. Combination of video and dataglove corpora

The third study is related to the use of numerical data provided by a capture device and synchronised with the video in annotation software (Crasborn et al. 2006).

A CyberGlove is used to capture hand movements and shapes. The corresponding data file is associated with an annotation document in the same way an audio or video file is. The tracks extracted from this file can be viewed as line plots, parallel to a horizontal time axis.

The benefit of an integrated visualisation of numeric data with video and annotations is twofold: the line plots assist accurate annotation tasks, and numerical values can be converted to annotation values. The resulting annotations can be exported to tab-delimited text file for further processing in a spreadsheet or in statistical analysis software.

Figure 5: Plots of numerical data in Elan.

3. Corpus for animation processing

SL corpora can also be used for animation processing of a virtual signer. Three main kinds of uses can be listed: use of video corpora to create 3D animations, use of motion capture corpora to create 3D animations, and use of motion capture corpora to create models for automatic computation of animations.

3.1. Video corpora for rotoscoping

Video corpora are used as models for the design of 3D animations with the rotoscoping technique. Rotoscoping was invented in 1920 for cartoons. The film was thrown on a table and a draftsman traced every image. We apply the same principle with modern software such as 3D Studio Max[tm] (Filhol et al. 2007). A computer graphist designer loads the reference corpus, uses it as a background, and synchronises following key frames timeline. A key frame corresponds to a 3D posture of the virtual signer's body. By using the different 3D views, the virtual signer's skeleton is exactly positioned on the image from the reference corpus (figure 6). The software needs as input only a selection of key frames and computes the intermediate positions. Key frames are set up when the movement changes; for example when a hand opens, key frames will be the closed, half-opened, and totally opened hand configuration. Finally, the computer graphist designer can modify the roundness of the gesture to obtain a perfectly realistic result, making

use of his artistic touch.

Figure 6: Rotoscoping technique.
WebSourd corpus.

3.2. Motion capture corpora for animation

Motion capture is an increasingly popular approach for synthesizing human motion. Much of the focus of research in this area has been on techniques for adapting captured data to new situations. Motion capture data can be reordered over time, similar motions can be interpolated, motion can be edited, and new motions can be generated by combining motions for individual limbs. Models of human motion can also be used to synthesize new motion.

An example of motion capture devices is shown in Figure 7. It shows that these kinds of devices can be intrusive in some case (left side of the figure), thus can harm to the representativeness of the captured data. Other motion capture devices are based on the use of intra-red cameras and little reflectors that are placed on specific locations on the body, providing less intrusive way to capture the gestures (right side of the figure).
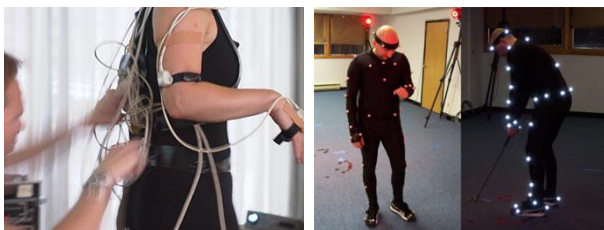

Figure 7: Motion capture devices.

Motion capture corpora are now also created to build SL animation (figure 8), picking isolated signs in a database and interpolating them together (Héloir et al. 2005).
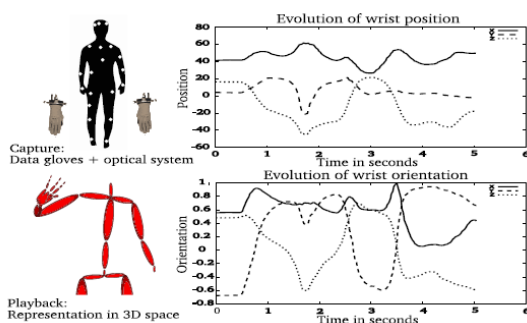

Figure 8: Rotoscoping technique.

3.3. Motion capture corpora for modelling

Another way to animate virtual signer is to generate animations automatically from symbolic descriptions based on linguistic models (Elliott et al. 2007). While progress has been done in this domain, virtual humans still lack naturalness in their gestural behaviour.

At LIMSI, we use a corpus of motion capture to build statistical models of the skeleton (Delorme 2010). From the corpus data, we extract information on the skeleton such as maximum and minimum angles for each joint, rest positions of the limbs, or the probability of the skeleton to be in a given position (figure 9).
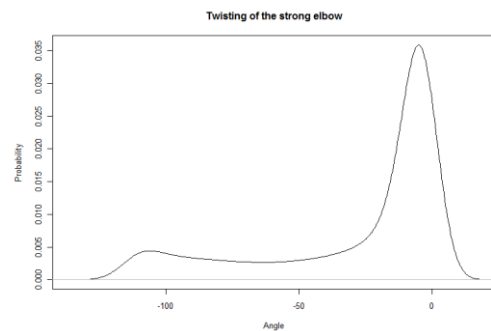

Figure 9: Statistical modelling of the elbow twisting.

These models, which have been computed from the corpora 3D data, will be used for automatic generation animation. The aim here is to automatically animate the virtual signer in a more human-like way than the current animation systems.

4. Corpus for model evaluation

Another use of corpora, though perhaps less obvious, is to help evaluate description models.
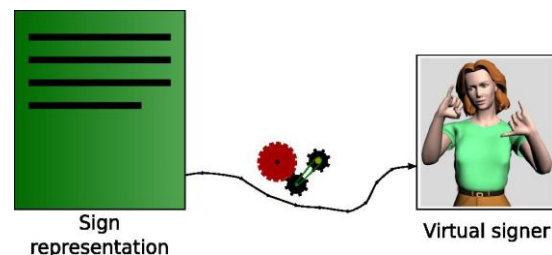

Figure 10: Description models, used for automatic generation.

Given a lexical representation model for signs, evaluating how well it covers the vocabulary of a language is not trivial, and it cannot be done without a decent amount of descriptions. A corpus of signs in their citation form (i.e. the context-free form that would constitute the lexical entry in a dictionary) helps to create entries for each of these signs and thereby to test the model's coverage.

LIMSI has devised a new geometric model for lexical representation (Filhol 2009), and has recently building a database of 1,500+ signs to test it. The model we have is

made not only to describe citation forms , but claims also to account for all context variations of a sign. A different type of corpus is needed then, as we need instantiated signs in various discourse contexts to include them all in the descriptions. This second stage of the description process will allow us to test the model's ability to include context sensitiveness and extend sign re-usability, which will also serve as an evaluation of its benefits –if any– over traditional approaches.

Figure 11: Extract from DicoLSF, the LIMSI's lexical sign database.

5. Conclusion

The way SL corpora are used, whether based on video or motion capture, have been extended these last years. This can be for language analysis, animation processing, or model evaluation.

The kind of data mainly used at this moment is video-based, but motion capture-based data are now used for various purposes.

At LIMSI, we plan also to use motion capture for analysis, like Malaia et al. (2008), and for evaluation. This implies to build a French Sign Language corpus using specific capture devices, as less intrusive as possible. Such a corpus, to be built with respect to socio-linguistic criteria ensuring that the corpus is representative for the study being considered, does not exist at present.

6. Acknowledgements

7. References

Braffort A., Choisier A., Collet C., Lejeune F. (2003), "Presentation of three French Sign Language corpora", 5th International Gesture Workshop on Gesture and Sign Language based Human-Computer Interaction (GW 2003).

Chételat-Pelé E., Braffort A.,Véronis J. (2008), "Annotation of Non Manual Gestures : Eyebrows movement description", 6th International Conference on Language Resources and Evaluation (LREC 2008), Workshop construction and exploitation of sign language corpora.

Crasborn O., Sloetjes H., Auer E. & Wittenburg P. (2006) Combining video and numeric data in the analysis of sign languages within the ELAN annotation software. In Chiara Vettori (ed.) Proceedings of the 2nd workshop on the representation and processing of sign languages: lexicographic matters and didactic scenarios. Paris: ELRA. Pp. 82-87.

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

Delorme M. (2010), "Sign Language Synthesis: Skeleton modelling for more realistic gestures", to be published in SIGACCESS Newsletter, "Accessibility and Computing".

Dreuw P., Rybach D., Deselaers T., Zahedi M., and Ney H. (2007). Speech recognition techniques for a sign language recognition system. In Interspeech, pages 2513–2516, Antwerp, Belgium, August.

Elliott, R., Glauert, J.R.W., Kennaway, J.R., Marshall, I., and Safar, E.,, "Linguistic modelling and lanuage-processing technologies for avatar-based sign lanuage presentation", Universal Access in the Information Society, vol. 6, no. 4, pp. 375-391, 2007.

Heloir A., Gibet S., Multon F., Courty N. Captured (2005), "Motion Data Processing for Real-Time Synthesis of Sign Language", in Gesture in Human-Computer Interaction and Simulation, Springer, 2005, 168--171

Filhol M., Braffort A., Bolot L. (2007), "Signing Avatar: Say hello to Elsi!", 7th International Gesture Workshop on Gesture in Human-Computer Interaction and Simulation (GW 2007).

Filhol M. (2009), "Zebedee: a lexical description model for sign language synthesis", LIMSI report 2009-08.

Malaia E., Borneman J., Wilbur R. (2008), "Analysis of ASL Motion Capture Data towards Identification of Verb Type", Semantics in Text Processing. (STEP 2008), Research in computational semantics, J. Bos and R. Delmonte Eds.

Liddell S. (2003). Grammar, gesture and meaning in American Sign Language. Cambridge Univ. Press, Cambridge.

Morrissey S. (2008) Data-driven machine translation for sign languages. PhD thesis, Dublin City University.

Segouat J. (2009), "A Study of Sign Language Coarticulation", SIGACCESS Newsletter, "Accessibility and Computing", pp. 31-38, issue 93, janvier 2009.

Thompson R., Emmorey K. Kluender R. (2006), The Relationship between Eye Gaze and Verb Agreement in American Sign Language: An Eye-tracking Study, in Natural Language & Linguistic Theory, vol 24/2, Springer.