Building 3D French Sign Language lexicon

Jérémie Segouat[(1)(2)], Annelies Braffort[(1)], Laurence Bolot[(1)], Annick Choisier[(1)], Michael Filhol[(1)], Cyril Verrecchia[(1)]
[(1)] LIMSI-CNRS
BP 133, 91 403 ORSAY Cedex, France
[(2)] WebSourd
99, Route d'Espagne, Bât. A, 31 100 Toulouse, France
E-mail: {forename.name}@limsi.fr

Abstract

We present a methodology to build 3D French sign language (LSF) corpus of lexicon. These signs will be used in various software dealing with signing avatar. One software we are developing is a display information system in a railway station, to provide as information to deaf travellers as the hearing can get. Another one is a Web dictionary of LSF lexicon with particular entry ways. The third software is a LSF translation on our laboratory's Web site to provide accessibility for deaf users. Our aim is to build a set of signs so that one sign can be used in each software listed. In order to create this 3D LSF corpus lexicon, we set up a new methodology. We propose four steps to ensure the quality of the result: selection of the participants, meetings to elaborate the lexicon, elaboration of the video corpus that will be used for the 3D conception, creation of the 3D corpus by means of 3D software. We suggest the participation of the end-users from the corpus conception, the participation of experts and specialists for the corpus content, and a multi-level evaluation (on technical, use, ergonomic and linguistic sides) of the result. Finally we present which step we have achieved in each of our software project

1. Introduction

Our study is in the field of automatic French Sign Language (LSF) generation and 3D virtual animation of a signing avatar. Our aim is to generate LSF utterances by concatenation of isolated signs, which would be performed by the signing avatar. We want to set up a 3D LSF sign corpus creation methodology together with its evaluation within various software.

The interest of this work is to propose a 3D LSF animation methodology which could be generic whatever the software that will use it: these 3D signs could be re-used regardless of the end-software. We are not saying that our methodology is generic for corpus creation, we just focus on corpora whose aim is to generate LSF.

With our approach, one sign can be used in different software, to produce a large amount of utterances (we use LSF lexicon signs, but eventually we could generate productive signs thanks to: Bolot (2006), Chételat-Pelé (2007), Filhol (2008).

In the next section (2) we introduce the notions of video and 3D SL corpora creation and how we plan to use the last one. Then, in section 3, we detail the four steps of our methodology: selection of the people who will participate to the corpus creation, conception of the corpus, building the video corpus, building the 3D corpus. In section 4 we present an evaluation methodology for the SL video and 3D corpus. Finally we present which step of our methodology we achieve for each software we are developing.

2. Video corpus and 3D corpus

There are two types of corpora: video and 3D. We elaborate the first one by filming a deaf person, following technical and linguistic specific criteria. These criteria differ from one goal to another. The second corpus is created by using 3D animation software. The animations are created by using a video reference of each sign.

Video corpora[1] are built for research studies, especially by researchers in linguistics. Because those corpora are created with a specific goal, it has consequences on the type (in vivo or in vitro), the technical characteristics (how many cameras, which shot, etc.), the selection of the panel (how many signers, which level of language, etc.), the linguistic content (isolated signs or full utterances, narration or dialog, etc.) of the corpus. There is no generic methodology for creating SL video corpora. There is a standardisation (IMDI project) but it applies to the metadata: Crasborn, Hanke (2003). In this project, he structure of the information about the corpus is normalized but not the conception.

Concerning 3D LSF corpus creation, as far as we know there is neither a corpus nor a methodology to create one.

At LIMSI, we have created several video corpora for previous projects with linguists, for example in: ARC-LSF and LS-COLIN[2] (Braffort & al., 2001) projects, and TALS[3] (2005). More recently, we have built 3D LSF corpora to be used with a signing avatar: the goal of the first one is to create a system providing information in a railway station, and another one is used to experiment concatenation of isolated signs to generate a LSF sequence giving the date of the day (Figure 1).

The sign corpora (video and 3D) we just talked about were created with different technical characteristics, goals and people in the film. Those creations give us experience

[1] http://www.sign-lang.uni-hamburg.de/BibWeb / http://www.bslcorpusproject.org http://www.bu.edu/asllrp/cslgr/
[2] http://www.irit.fr/LS-COLIN
[3] http://tals.limsi.fr/

better to choose our methodology depending on the end-use of the corpus.

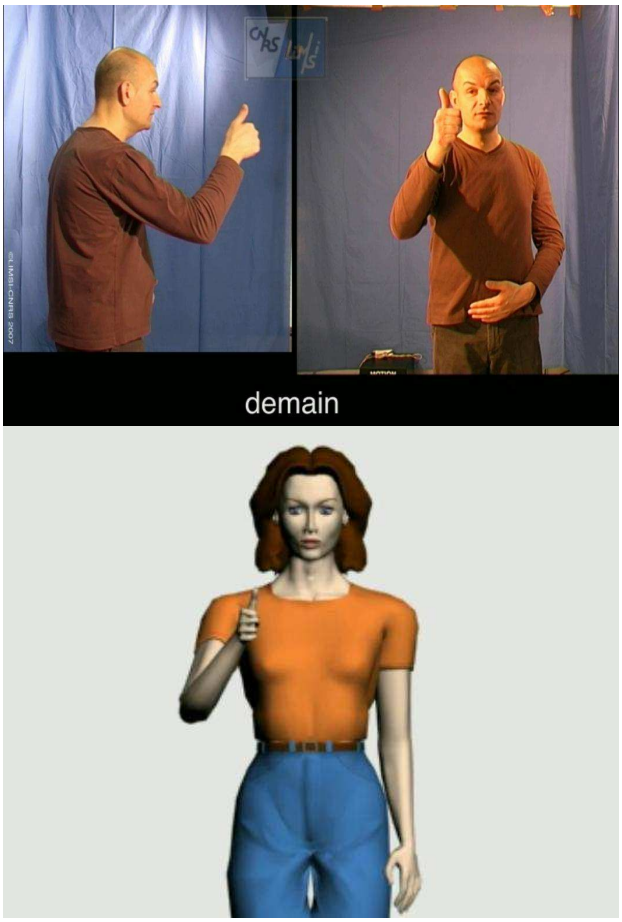

Figure 1: Video and 3D corpus

We intend to use our 3D LSF sign corpus in three types of different software:

- a railway station information system. As hearing people have access to the information given by a voice generated system, we aim to display a signing avatar on a screen to inform deaf people about general events (keeping an eye on luggage, etc.) and more specific ones (the delay of a train or a change in the platform number where a train will arrive, etc.);
- a bilingual LSF / French dictionary on the Web. We would like to provide Web users with a 3D sign dictionary to avoid the video drawbacks (non anonymous signer, impossible modifications, etc.) and to structure their display with several ways that are not yet implemented in other Web LSF dictionary (Moreau, 2007): by parameters, or by more complex properties like symmetry, re-localisation, etc. (Filhol, 2007);
- an ECA (Embodied Conversational Agents) on the Web site of our laboratory. It will interactively translate or explain in a LSF way specific words and concepts of our laboratory's research fields.

The software listed above is of different types in terms of graphic user interface (isolated sign or utterances), target users (in a station, on a web site), and cognitive context: the goal of the users is different if they want information about a train, or about lexicon, or if they want a translation or an explanation on a word or a concept. Nevertheless, we want to use the same 3D signs corpus (numbers, for example, would be used in the three projects). In order to achieve this goal, we propose a methodology detailed in the next section.

3. Methodology

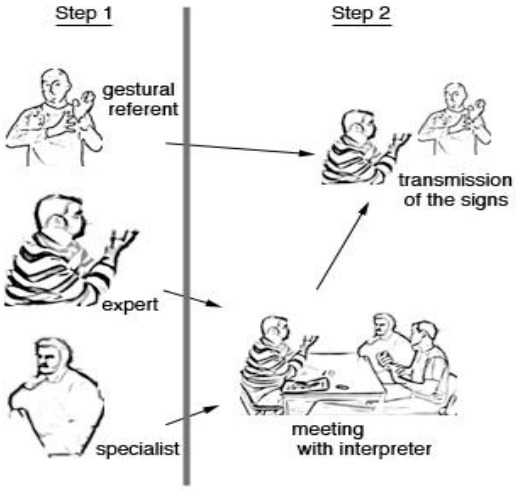

Figure 2: methodology 1/2

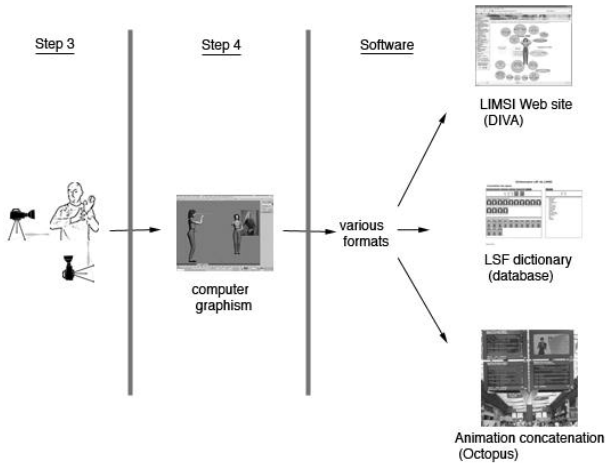

Figure 3: methodology 2/2

Our goal is to have a generic 3D LSF sign corpus. To build it, we need a video reference of each sign. The whole process is composed of four steps, which are represented in a graphical form in the figures 2 (steps 1 and 2) and 3 (steps 3 and 4).

1) The first step is to define who will be our « gestural referent ». In other words, how we select the person who will be filmed signing every LSF sign we need. This person must be a deaf and skilful user of LSF, expressing clearly and easily. The person is proposed by our team and validated by our computer graphist. The team's choice is

made with deaf and fluent people in LSF, and the computer graphist validates the choice depending on how it matches with the work he will have to do with the video. At this moment, we have two gestural referents, so we have to keep in mind that we will have to evaluate combining 3D signs build with video from both referent in terms of quality perception (and comprehension) of the message by the end-user. There remains a little difference in style even if the computer graphist, when creating 3D signs, makes some smoothening (see step 4). During a previous project where information messages were displayed in a railway station, end-users (travellers) easily recognized the person who was chosen to be our gestural referent and whose video signs were used to create the 3D animation. Comparing this previous project and the present one, the difference is that we now build a corpus of isolated signs, and not an utterance corpus. Combining isolated signs, even if they were created from a unique referent, with context and co-articulation influences, should erase this phenomenon. We'll see, with the first evaluation results if this hypothesis is confirmed.

At the same time, we contact « specialists » of each field we want to get lexicon of.

For the railway station project, lead by the SNCF (French national railway company), we have a partnership with WebSourd. We have to create messages similar to the audio announcements. For instance, « TGV number 1234 with destination Rennes, stopping in Le Mans and Laval, leaves at 8h56, from platform number 19 », or « due to an incident at level crossing, train number 35750, normally arriving at 13h01, will be delayed about 15 minutes ». SNCF provide all sentences that have to be displayed with the LSF 3D signing avatar. Then WebSourd provide a translation in LSF. We then dispose of the video corpus. In this project, the filmed gestural referent's is a LSF/French translator. Therefore, he is an expert of the two languages, working with his team to find the most accurate translation. He is used to be filmed because of his occupation, so he fully verifies our criteria to be a gestural referent. The video corpus is built by filming each sign isolated and in a sentence in step 3.

For the dictionary and the laboratory's Web site, in addition to the previous criteria, we needed a person with a significant general knowledge to express signs from a paper dictionary (IVT, containing signs of the everyday life) and a simple lexicon of our laboratory's Web pages (for the specific lexicon of the pages, we call on an « expert »). Consequently, we work with a deaf actor.

Still at the same time, we look for « experts » : deaf persons, skilful in a field for which we want some translations and/or explanations in LSF (if possible, all fields of our laboratory) or from connected fields (because it is not easy to find deaf people working in the same field as our laboratory).

2) We then begin the second step which is the conception of the video corpus. If the gestural referent has enough skills in the knowledge field he is going to express signs, we shoot directly, in step 3(of course, the referent prepare the signs before being shot). Otherwise, we call the expert and the specialist. The expert meets the specialist better to understand all the words or concepts he is not sure of. This meeting takes place with a LSF - French interpreter, so there could be an influence in some ways. If there is a discussion on a particular French word but there is no word-to-sign translation, the first solution used by the interpreter will be a circumlocution based on choices due to the translation principles. This influence could be erased partially if there is a discussion between the expert and the specialist: the expert can reformulate what he has understood and the specialist can validate or be more accurate rephrase to. Even is the influence is minimal, it exists and we have to keep it in mind while evaluating the users' answers on the evaluation time. The meeting is recorded on video tapes, stored for the expert, but also to take into account the influence of the interpreter's translation while evaluating the end-users' answers about the quality of the translation. Lastly, the expert meets the gestural referent to give him every sign that will be shot for the video corpus.

3) The third step is to film the video corpus. Following the computer graphist recommendations, we chose a double shot: side and front medium shots. Those two shots are edited together in a single video file given to our computer graphist.

4) He conducts the fourth and last step which is the 3D signs corpus creation based on the videos corpus. The creation is made by copying key frames of the video manually, with 3DSmax™. Those 3D animations are then converted in various formats, depending on which software will use them (Octopus[4], DIVA[5], database). We finally have our 3D LSF sign corpus.

4. Evaluation

As far as we know, there are few studies on evaluation of video and 3D corpora evaluation. Huenerfauth (2007) proposes an ASL generation evaluation methodology, but his work is based upon classifier predicates (CP) generation, while ours at present focus on generating standard lexicon. CP's generation is evaluated through the user's choice of an animated sequence that translates what was displayed in ASL. The evaluation that Hunerfauth suggests is not immediately usable for our work, because his goal was to evaluate the rightness of the CP's choice and not lexicon animation. Nevertheless we will follow the generally accepted principles (like Huenerfauth does) of using questionnaires to get the users' opinions.

During the evaluation, we will gather information about the software and the 3D corpus. The evaluations will be different depending on the software's characteristics. For the information system in railway stations, we will propose a questionnaire in the station (a face to face interview in LSF) and a video questionnaire on a Web site. The answers will give us critics on ergonomic, technical, linguistic and use aspects. The panel of the users should be

[4] http://www.limsi.fr/Individu/bolot/octopus.html

[5] http://www.limsi.fr/Individu/jps/online/diva/divahome/index.html

composed of deaf users of the railway network and deaf specialists of LSF and information and communication technology (ICT). For the dictionary and the accessibility of our Web site, users will give their opinion by written or video email, and a video questionnaire will be proposed on the two Web sites. Questions will focus on ergonomic, relevance of the proposal entries and correctness of the signs (and of the translation), and, in particular for the Web site accessibility project, on the comprehensibility of the signing avatar's LSF and its overall ergonomic quality. We will ask users via a questionnaire on the Web sites and we will invite LSF and ICT specialists to test the system.

5. Conclusion

We want to dispose of a LSF 3D isolated signs corpus to be used within various software. We suggest a 3D sign creation methodology from a video corpus. The video corpus itself is built following a particular methodology: selection of a deaf gestural referent, possible selection of a deaf expert and a specialist, and selection of signs by the gestural referent or the expert after meeting with the specialist. This process is interesting because we will lastly have a set of 3D signs that every software can use according to the needs. At the beginning we will also have to propose validated signs by the gestural referent which rely on an existing paper dictionary, or by the expert. The validation will be iterated: signs will be displayed to the public, it will evaluate it, and we will modify the signs according to the public feedbacks, and will re-propose them to the public, and so on). This validation process will be the same for the signs and for the 3D creation methodology proposed.

The methodology we set up should guarantee us a significant relevance, thanks to the discussion between deaf experts and specialists of the field, and a good quality of the video signs, and 3D signs too, thanks to the selection criteria of the gestural referent. The evaluation step should guarantee maximal feedback about our methodology validation all along the four steps. What will be evaluated is the gestural referent selection (his style) for the first step, translation choices by the expert for the second, the quality of the video corpus for the third, and the quality of the 3D corpus for the last step. Of course, we will also evaluate the overall software.

We are at the first step of our methodology concerning the information system, at the second step for the accessibility of our laboratory's Web site, and we already have passed the third step for our dictionary project. The beginning evaluation step will confirm our methodological choices, and will give us ways to make it better.

6. References

Bolot, L. Braffort, A. Filhol, M. (2006). Elsi balbutie ! Vers une plateforme d'animation d'avatar signant. *Actes du 2ème Workshop Francophone sur les Agents Conversationnels Animés 2006 : WACA 2006*, Toulouse.

Braffort, A. Choisier, A. Collet, C. Cuxac, C. Dalle, P. Fuselier, I. Gherbi, R. Jausions, G. Jirou, G. Lejeune, F. Lenseigne, B. Monteillard, N. Risler, A. Sallandre, M.-A. (2001). Projet LS-COLIN. Quel outil de notation pour quelle analyse de la LS ?. *Actes de Journée Recherches sur la langues des signes*, Toulouse.

Chételat-Pelé, E. Braffort, A. Véronis, J. (2007). Mise en place d'une méthodologie pour l'annotation de gestes non manuels. *Actes de TALS 2007. Traitement Automatique des Langues des Signes 2007 : atelier de Traitement Automatique des Langues Naturelles 2007*, Toulouse.

Crassborn, O. Hanke, T. (2003). Additions to the IMDI metadata set for sign language corpora. ECHO Project KUN, Taalwetenschap.

Filhol, M. Braffort, A. (2008). Description lexicale des signes. Intérêts linguistiques d'un modèle géométrique à dépendances. Revue *Traitement Automatique des Langues*, Volume 48 Numéro 2 : *Modélisation et traitement des langues des signes* (à paraître).

Huenerfauth, M. Zhou, L. Gu, E. Allbeck, J. (2007). Evaluating American Sign Language Generation Through the Participation of Native ASL Signers. *Ninth International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS-200,*. Tempe, Arizona, USA.

Moreau, C. Tranchant, C. and al., (2007). Modèle de recherche de signes standard en langue des signes française, *Traitement Automatique des Langues des Signes TALS 2007, TALN'07*, Toulouse, 5-8 juin 2007.