

Toward Modeling Sign Language Coarticulation

J r mie Segouat^{1,2}, Annelies Braffort²

¹ WebSourd, 99 route d'Espagne, 31100 Toulouse, FRANCE

¹ LIMSI-CNRS, BP 133 F-91403 Orsay Cedex, FRANCE
jeremie.segouat@limsi.fr

Keywords: Sign Language Processing, Annotation, Coarticulation Modeling.

1 Introduction

Our researches focus on Sign Language (SL) gestures, which can be manual or non-manual (like facial expressions). This abstract presents the study we conduct in the field of French SL (LSF) processing. We intend to propose a coarticulation model of signs of LSF, thus we need accurate analysis of both isolated and in utterance signs. We follow a several step methodology: video corpus creation, annotation, statistics computation, and model development.

2 Modeling: from definition to elaboration

Our aim is to set up a coarticulation model of LSF. First of all we have to precise what "coarticulation" means. Studies on coarticulation have been made in speech and gesture research fields, and also in SL recognition [1] but not in SL animation. After collecting all the possible definitions, we have chosen to define coarticulation as modification (either addition or deletion) of signs when using them in utterances compared to when using them isolated.

To achieve our goal, we firstly create a video corpus of signs. Then, we annotate the videos, thanks to ANVIL [2], video annotation software. Third step is analyzing the annotations by computing statistics, and lastly we propose a model based upon the statistical results.

The video corpus is composed of both isolated and in utterances signs. For instance, we record isolated signs like "5", "6", "A", "B", and utterances "the train will leave at 5 p.m. from platform A", "the train will leave at 5 p.m. from platform B", and "the train will leave at 6 p.m. from platform B". This allows us to compare each isolated sign to how it is performed within utterances. In Figure 1 we show how different a sign performance could be, depending on if it is isolated or in utterance. It is obvious that modifications are not so simple: in our example we see that both hand and facial expression variations have to be taken into account.

We are looking for modifications between and inside signs. Thus we need to describe accurately what occurs before, during and after a sign is performed. This is done thanks to the Johnson & Liddell descriptive model [3], enhanced by another study [4]. We annotate both manual and non-manual features following the descriptive model. We also annotate from a "discourse analysis" point of view, because it may have some overall influence on parts or all of utterance.

Then, we run statistics on these annotations in order to find correlations between annotated elements. For instance, we look for repetition, simultaneity, and inclusion of elements within others, both inside a sign (isolated and in utterance) and between two signs in utterance.

Our model is lastly built upon our statistical results. It is a set of rules that will be used to coarticulate signs when they are used in utterances. These rules are built following two linguistic points of view. On one hand we use a descriptive aspect of what occurs phonetically in utterances, on another hand we use a semantic aspect to take as a whole the context of the utterances.



Fig. 1. On the left side, the sign [10]_{LSF} is performed in an utterance; on the right side the same sign is performed in isolation.

3 Conclusion

Our goal is to build a model of coarticulation in LS [5]. In order to reach our goal we create a video corpus, annotate it, analyze the statistical results based on annotations, and elaborate our model. At present, we are annotating the corpus, and in the following weeks we will be able to analyze these annotations. Next step will be to propose a first coarticulation model for SL. After evaluating this first attempt, we will improve it, annotate a bigger corpus, and propose an enhanced model.

All versions of the model will be evaluated both in laboratory, and in concrete software (a display information system in railway stations).

References

1. C. Vogler, "American Sign Language Recognition: Reducing the Complexity of the Task with Phoneme-Based Modeling and Parallel Hidden Markov Models", Ph.D. Thesis, Department of Computer and Information Science, University of Pennsylvania, 2002.
2. M. Kipp, "Anvil - A Generic Annotation Tool for Multimodal Dialogue", 7th European Conference on Speech Communication and Technology, Aalborg, 2001, pp. 1367-1370.
3. R.E. Johnson, and K.S. Liddell, "Sign language phonetics: architecture and description", In preparation.
4. M. Filhol, "A descriptive model of signs for Sign Language processing", Ph.D. Thesis, LIMSI-CNRS, University of Paris Sud, 2008.
5. J. Segouat, "A study of signs language coarticulation", ASSETS 2008, Tenth International ACM SIGACCESS Conference on Computers and Accessibility, Halifax, Canada, 2008.