What constraints for representing multilinearity in Sign language?

Michael Filhol, Annelies Braffort

What constraints for representing multilinearity in Sign language?

Michael Filhol, Annelies Braffort

LIMSI-CNRS, B.P. 133, F-91403 Orsay cx, France
{michael.filhol,annelies.braffort}@limsi.fr

Abstract. In this paper, we present our current investigations on one of the main challenges in Sign Language modelling: multilinearity. We explain the way in which we extract grammatical rules from Sign Language corpus, and how we represent them with constraints. Two kinds are needed: time synchronisation and geometric constraints.

1 Introduction on Sign languages

Sign languages (SLs) are visual-gestural natural languages used by the deaf communities. SLs are less-resourced languages and have no writing system. In other words, there are very few reference books describing these languages (grammar rules, etc.), a limited number of dictionaries, and mainly small-sized corpora. Moreover, SL linguistics and processing research is quite recent, and a proper formal description system of the rules governing SL is still to be proposed.

In view of building SL processing systems (recognition, generation, translation...), we need a database of representations for SL elements such as lexical units or grammatical rules. Formal knowledge on these still being in its infancy, this requires thorough investigation beforehand, based on annotation and analysis of SL corpora.

The main challenges in SL modelling are: the relevant use of space in SL discourse, the iconicity that is present at various levels (iconic morphemes, spatial agreement...), and the multilinearity of the language. This paper gives an overview of our current investigations on this last property, started during the European project Dicta-Sign[1]. Section 2 explains how we analyse corpora to build our rules, and section 3 presents our use of constraints to represent them.

2 Building rules by linking form to function

The first important fact about SL is that **all body articulators from the waist up**, whether manual, non-manual, skeletal or muscular, can be relevant as they may all—though they do not always—carry meaning. Therefore, all visible activity in a Sign production must be included in the study.

[1] http://www.dictasign.eu

We have observed and partially annotated the French part of the DictaSign corpus [10] and the older LS-Colin corpus [2]. With as few articulators as eyebrows, eye-gaze and both left and right hand activity (and yet the signing body involving many more), it is already clear that **articulators do not align** in any trivial way. That is the second important specificity of SL to deal with.

With these two SL characteristics in mind, we proceed with the corpus study to find systematic links between:

– *surface form features*, i.e. observable/annotated movements or positions of body articulators like a cheek puff, directed eyegaze, hand location change or shoulder line rotation;
– and *linguistic function*, i.e. the possibly language-dependant interpretation of the form, taken to be part of the signer's discourse intent.

Funtions can be identified on all layers of the language; none is linked to a specific one. Any linguistic operation purposely produced may be called a function, whether it inclines to a morphemic, lexical, syntactic or even discourse level. For example, turning a clause into an interrogative, using a known lexical sign, adding modality to a verb or outlining a full speech with intermediate titles are all possible functions. The point for our study is now to find features of the production form that are consistantly used over many occurrences of a linguistic function. Factoring out those form *invariants* raises a candidate set of necessary features for a given function. Describing this set of features under a label for the function raises what we call a *production rule* for the function.

For example, it is already well-documented that in LSF, space locations can be activated by pointing at the target in the signing space with the strong hand, immediately after directing the gaze towards the same target. Other elements are sometimes held on facial articulators or performed by the weak hand, but they do not occur consistantly while the eyegaze-then-point structure is invariant.

Formally describing the invariant above will create a production rule which can be labelled "point and activate locus". We explain how we handle those descriptions in the next section.

3 Constraints to represent production rules

There are two problems here to address. One is time synchronisation of parallel activity on a multi-linear signing score, in our example the simultaneity of gaze and index. The second is articulation of body articulators in space, e.g. the hand shape for the pointing gesture and the common direction for both eyegaze and index.

The next section explains the limits in existing models; the following two present the approach we use at LIMSI.

3.1 Traditional models

Most models in traditional Sign linguistics do not account for articulator synchronisation issues as fundamental elements of the SL discourse. They rather define elementary movements and/or sign units that can be concatenated [12, 11]. Some models [8] do promote time to a more essential place in their descriptions. Liddell and Johnson have shown that signs reveal (rather short) periods when all "features" of the hands align to form intentional **postures**, between which **transitions** take place to change from one established posture to the next (see fig. 1a further down). However, all of these models have only really been used (and even designed in the first place) for lexical units of SL, primarily even for "fully lexical signs" (i.e. non depicting). Though, SLs generally enrol more body articulators and even provide with productive ways of creating unlisted iconic units on the fly [3, 7].

Unfortunately, not much exists on SL formal modelling on the clause/sentence level. The most relevant progress is the ViSiCAST[2] and eSign projects and their generative HPSG framework [9]. The basic timing unit in all generative approaches is the lexical item at the bottom of the syntax tree. Thus, any phonetic feature of the body must be borne by a lexical unit composing the clause and the productions burn down to a single tier carrying a lexical sequence. Articulator synchronisation is then only possible using the lexical boundaries.

A first account for multi-track description was given by the "P/C" formalism [6]. It defines two ways of conjoining parts of a signing activity: *partitioning* for parallel specification and *constituting* for sequence. But we argue [5] that it does not allow for abstract description of parts that overlap in time. Section 3.2 presents our approach to tackle this problem with free sets of constraints.

In addition to the timing issues, another question raised is how to specify the signing activity happening at a given time. The problem with L&J's representation is that every moment in time requires the same fixed set of parametric values, numbered a, b, c... in fig. 1a. But similarly to time constraints, we want this new specification to be flexible hence allow for under-specification if it is needed. Section 3.3 below deals with this issue.

3.2 Time synchronising constraints

While L&J produce good evidence that much of the signing activity does synchronise with postures, observation reveals that many constructs fall out of the scheme and require a more complex synchronisation technique. Many productions break the set of body articulators into subsets that act more or less independently, each in its own active period (a time interval, TI henceforth). Sometimes the full face acts as a group while the hands are moving; other times a few facial parts take on a different linguistic role and de-synchronise from rest of the face and hands; hands do not always work together; etc. Our proposition is to combine those two sync schemes into a same language for a better coverage of SL linguistic structures. We propose the **AZee** language, combination of the following two.

Zebedee is a language designed to specify the posture–transition patterns [4]. A description in Zebedee (or *zebedescription*) specifies:

[2] http://www.visicast.cmp.uea.ac.uk

1. an alternation of “key posture” (KP) and transition units (the horizontal axis in figure 1b below;
2. the articulation that takes place in each of those units (this is further explained in the next section).

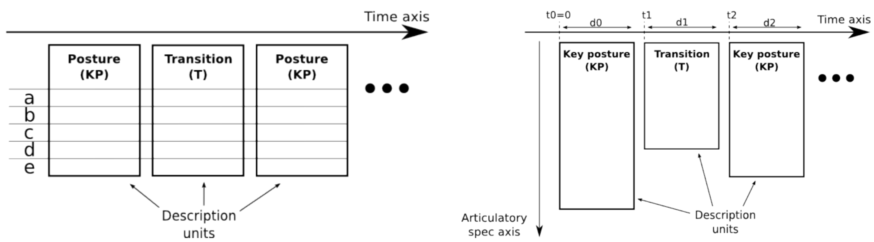

Fig. 1. (a) Liddell and Johnson’s phonetic sequence with a fixed number of parameters; (b) Timing units in Zebedee: each unit is described with a variable number of elements

Azalee is a language designed to specify the free-hand (i.e. not necessarily aligned) synchronisation patterns for sets of separate time intervals (the TIs) on a timeline. Each Azalee specification (or *azalisting*) is a container for the two following parts, together specifying the intended rule fully:

1. all (usually overlapping) TIs are arranged on the timeline following temporal precedence constraints;
2. each TI specifies the signing activity taking place within.

In part (1), TIs are time-arranged with a *minimal* yet *sufficient* set of constraints in order not to over-specify the surface production and allow for any needed flexibility. The reason for preferring minimality for these sets is that we observe a lot of nested structures, all acting on the body articulators simultanously. Any superfluous constraint on an articulator in a rule will prevent other rules to use it freely, which may result in unsatisfactory productions.

The temporal precedence constraints can be of any form, whether strict or flexible, of precedence or duration, over full TIs using Allen logic operators [1] or over single boundaries. In part (2), a TI can either make use of a nested Azalee description, or use Zebedee for a simpler sync pattern.

Figure 2 illustrates our example rule. Its Azalee pattern (outer box) is composed of 2 TIs, ‘eye gaze’ and ‘pointing sign’, which are zebedescriptions (KP–T alternation patterns). This flexible combination of two nestable timing patterns allows to account for complex synchronisation between body articulators.

3.3 Articulatory constraints

Fixed sets of parameters or features are not satisfactory to describe articulatory postures. As we have explained:

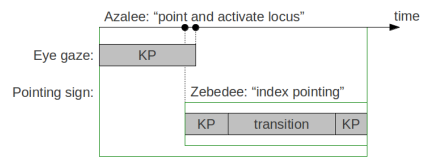

Fig. 2. AZee: an example of Azalee and Zebedee combination

– on the one hand, articulators are not all linguistically relevant all the time, thus including all of them in a fixed-set model would lead to a lot of over-specification;
– on the other hand however, any articulator of the upper-part of the body can carry meaning at some point, thus any exclusion from the fixed set leads to possible under-specification.

To avoid both problems, we choose sets of necessary and sufficient constraints to specify both as much and as little as needed at any moment in time. In other words recalling the minimality of the sets of time constraints, we aim at minimal sets for the articulatory constraints. The constraints are mainly of two types: orientation (to align bones along vectors) and placement (to bring a body location to a point in space or in contact with another).

For example, throughout the pointing gesture, we need to extend the index, which takes at least the following constraint, literally aligning the proximal index bone of the strong hand (side `S` below) in the same direction as the second:

```
Orient DIR!index(S,1) along DIR!index(S,2)
```

Placing the hand—the palm site (`@PA`) of the strong hand—in front of the body (forward fron the sternum site `@ST`) for the initial key posture will be written:

```
Place @WR(S) at @ST + <FWD | medium>
```

Each timing unit holds a conjunction of such constraints, and the resulting set is as large as necessary and as small as sufficient. Figure 1b illustrates this approach and contrasts with previous figure 1a. The vertical depth needed to specify body articulation in a zebedescription may vary in time, according to the number of articulatory constraints. Note that they are fundamentally geometric; all typical geometric transformations (translations, symmetries, etc.) are supported.

Also, expressions can contain context references enclosed in square brackets to account for significant (often iconic) context variations. For our example above, the distal bone of the pointing finger `!index(S,3)` must be oriented to point at the target, which is a context dependency. The constraint will be something like[3]:

```
Orient DIR!index(S,3) along <ORIGIN!index(S,3), [target]>
```

[3] Operator `<..., ...>` creates a vector from two geometric points.

4 Prospects

We have presented the issues in formalising systematic links between form and function. Our methodology, based on SL corpus analysis, is applied regardless of the level, from the sub-lexical level to discourse structure. The formalisation of the production rules uses the AZee model, which allows to express Sign articulator synchronisation with time constraints, and body postures and movement with geometric constraints.

While the Zebedee part of this model is already well assessed (2,000+ LSF signs described), we have so far only really explored a dozen Azalee structures. We intend to continue this work to increase this number. A thorough evaluation will then be a starting point for implementation of AZee rules in Sign synthesis software applications, a longer-term prospect being text-to-SL translation.

In parallel and in order to collect 3d/spatial data, we will have to collect a new kind of SL corpus, built using a motion capture system. This will allow us to conduct similar studies on the spatial organisation of SL and then provide with a finer representation of the signing space and its use in SL discourse.

References

1. Allen, J. F.: Maintaining Knowledge about Temporal Intervals. ACM 26:11, pp. 832–843 (1983)
2. Braffort, A., Chtelat-Pel, .: Analysis and Description of Blinking in French Sign Language for Automatic Generation, in Gesture and Sign Language in Human-Computer Interaction and Embodied Communication, LNCS/LNAI vol. 7206, Springer (2012, tbp)
3. Cuxac, C.: Langue des signes franaise, les voies de l'iconicit. Ophrys, Paris (2000)
4. Filhol, M.: Zebedee: a lexical description model for Sign Language synthesis. LIMSI-CNRS, Orsay (2009)
5. Filhol, M.: Combining two synchronisation methods in a linguistic model to describe Sign Language, in Gesture and Sign Language in Human-Computer Interaction and Embodied Communication, LNCS/LNAI vol. 7206, Springer (2012, tbp)
6. Huenerfauth, M.: Generating American Sign Language classifier predicates for English-to-ASL machine translation. PhD thesis, University of Pennsylvania (2006)
7. Johnston, T., Schembri, A.: Australian Sign Language (Auslan): An Introduction to Sign Language Linguistics. Cambridge University Press (2007)
8. Liddell, S. K., Johnson, R. E.: American Sign Language, the phonological base. Sign Language studies 64, pp. 195–278, Cambridge University press (1989)
9. Marshall, I., Sfr, .: Sign Language generation in an ALE HPSG, in Proceedings of the HPSG04 conference, Leuven, Belgium (2004)
10. Matthes, S., Hanke, T., Storz, J., Efthimiou, E., Dimiou,, N., Karioris, P., Braffort, A., Choisier, A., Pelhate, J., Safar, E.: Elicitation tasks and materials designed for DictaSign's multi-lingual corpus. LREC, Corpora and Sign Language technologies, pp. 158–163 (2010)
11. Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Henning, J.: HamNoSys version 2.0, an introductory guide. International studies on Sign Language communication of the Deaf, Signum press, Hamburg, vol. 5 (1989)
12. Stokoe, W.: Sign Language structure: an outline of the visual communication system of the American Deaf Studies. Linguistics, occasional papers, vol. 8 (1960)