Non-manual features: the right to indifference

Michael Filhol, Mohamed Nassime Hadjadj, Annick Choisier

LIMSI–CNRS
B.P. 133, 91403 Orsay cedex, France
Email: {michael.filhol,hadjadj,annick.choisier}@limsi.fr

Abstract

This paper discusses the way Sign Language can be described with a global account of the visual channel, not separating manual articulators in any way. In a first section section it shows that non-manuals are often either ignored in favour of manual focus, or included but given roles that are mostly different from the mainly hand-assigned lexical role. A second section describes the AZee model as a tool to describe Sign Language productions without assuming any separation, neither between articulators nor between grammatical roles. We conclude by giving a full AZee description for one of the several examples populating the paper.

Keywords: Sign Language modelling, non-manual features, synchronisation, AZee

1. Unjustified hand focus

1.1. Why ignore articulators?

Pretty much since the beginning of its description, whether naïve or scientific, SL has been "a way to speak with the hands". Initiated with Bébian (1825), established by Stokoe (1960) and completed by Battison (1978), the idea of formal Sign phonology through the description of manual parameters is still the most widely accepted way of describing signs. The number of technical projects involving SL, whether for its synthesis with signing avatars or its recognition with all sorts of devices (video tracking, Kinect, gloves, ring/bracelet sensors), unquestionably regard manual activity as the centre of all signed productions and the key to any underlying structure.

The parametric model eventually integrated an additional "facial expression" parameter, justified even by minimally contrasting lexical pairs such as "skin" vs. "racist" in LSF. But one must admit that it is usually discarded from lexicon descriptions, and occurrences of facial expression change over a lexical unit is often labelled grammatical or prosodic, i.e. almost off the limits of linguistic description, while a manual change in location, orientation or movement will likely be syntactically analysed.

For example, we have recently published the result of an LSF corpus study on event precedence and duration (Filhol, 2013). In this study, all sequences of two events separated by a period longer than 10 days involved the form photographed in fig. 1 and described as (r3) in the cited paper. It is close to the sign glossed "until now" in LSF picture dictionaries (fig. 2), at least in meaning but enough in form also to be used in annotation tasks. However, all occurrences in our study differ in the same way to what the drawing in fig. 2 suggests: the movement is of the linear type (not accelerated/ballistic), the fingers wiggle, the head rotates to the active hand's side, the eyes blink just before the gaze turns to the active side, the torso leans to the opposite side, and the synchronisation of all these features is consistent, etc. Why does no lexical description include those features on the same level as the manual gesture? Parametric descriptions do not even allow the torso tilt, but what makes the feature less lexical then the manual part, whereas we observe the former on every occurrence of the latter?

Figure 1: Snapshot of a form used for periods lasting over 10 days

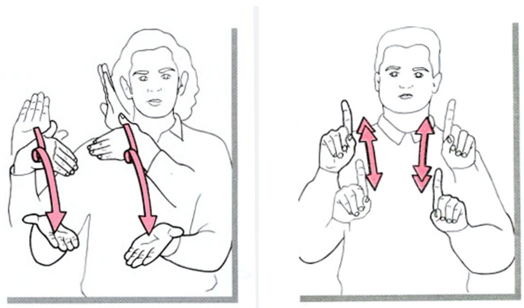

Figure 2: LSF picture dictionary images for "until now" (left) and "party"/"national day" (right)

Sometimes observable movements can indeed be side effects of other relevant body gestures, but this does not generalise to, say, the eye blink or the head rotation in our example. Conversely, parametric descriptions give a hand orientation in the LSF sign for "party"/"national day" like

palms face to face or away from body, whereas all variations occur, the only constraints invariably observed being the articulatory limits at the wrist. Also, should the coarse "facial expression" parameter be used to account for a facial detail such as an eye blink?
With no intention of denying the obvious articulation of hands in signed discourse, we do think the imbalance in focus between their part and that of other gestures should be questioned, manual preference not being justified and leading to flawed observations.

1.2. Why assign roles to articulators?

Not all published work fully discards non-manuals. Different studies exist, and a lot of them conclude assigning syntactic roles to observed articulators: manual placements on the left-right axis for absolute time anchors vs. on the sagittal axis for relative, eye gaze to switch between frozen and depiction mode (Cuxac, 2000), eyebrows combined with head tilts serve as interrogative markers (Baker-Shenk, 1985; Hickok et al., 1996), shoulder line rotation for reported turn taking (Moody et al., 1986), wide eye opening for the adverbial function of quantity (Vergé, 2001), etc.
In our study cited further up, all expressions of event duration, if exceeding 10 days, were found (r4) to involve the same form as fig. 1, including all articulations described but with an additional non-manual feature of semi-closed eyelids, which we note "el:semi-cl". Using the cited rule numbering system where (r2) is the chronological sequence, it is that only feature that differentiates the signed sequence (*event1*, fig. 1 with *duration*, *event2*) between the two meanings below:

- *event1* and *event2* are separated by the given *duration*: r2(*event1*, r3(*duration*), *event2*);
- *event1* is followed by *event2* lasting the given *duration*: r2(*event1*, r4(*duration*, *event2*)).

When the time period is under 10 days, the differences between event separation time (r1) and event duration time (r5) are:

- a change in manual activity—rule (r1) making use of a dictionary sign glossed "immediately followed by, your turn, consequence" whereas (r5) uses one glossed "duration, to last";
- (r1) uses eye gaze whereas (r5) does not;
- (r1) brings the chin forward whereas (r5) brings it up a little;
- (r5) uses the el:semi-cl feature whereas (r1) does not.

The change in manual activity in the case of shorter periods will unquestionably lead the traditional approach to call a lexical difference, optionally commenting on the non-manual features. But the case of longer periods less trivially allows overlooking the non-manual feature. Is the whole form (fig. 1+el:semi-cl) a different lexical item to fig. 1 alone, and to be glossed something like "during"? Or should we assign el:semi-cl the grammatical role of, say, denoting simultaneity of the period and the event to be signed afterwards? Indeed both forms of duration vs. separation use the same feature. Then what about the other non-manuals involved in the latter?

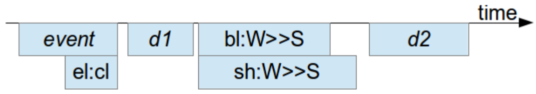

Figure 3: Form diagram for periods between two time boundaries, excl. enunciation time

Figure 4: Snapshot of a form used for periods between two time boundaries, excl. enunciation time

What is more, we have furthered our study since the cited publication, analysing the case of event durations defined by two time boundaries, as would be in English with the expression "from 1905 to 1914". We have found the two additional properties below:

- if the duration is disconnected from the present (enunciation) time and however long it lasts, the form is invariably that sketched in figure 3 (see picture in fig. 4), where:
  - *event*, *d1* and *d2* are the arguments denoting the event and the start and end dates of the period, respectively;
  - "el:cl" stands for an eye blink;
  - "sh:W>>S" is the 'J'-shaped strong hand lateral movement from weak to strong side;
  - "bl:W>>S" is the body movement leaning from weak to strong side, simultaneous to that of the strong hand;
- if the starting boundary is the enunciation time, e.g. "until Tuesday", the form used is that of figure 5 (see picture in fig. 6), where:
  - *event* and *until* are the two arguments, one being placed in either of two time positions;
  - "eg:s-sp" stands for an eye gaze directed to the signing space where the hands are placed;
  - "sh" and "wh" respectively stand for the strong and weak hands;
  - "hd:rot-dwn" is a small head rotation bringing the chin down.

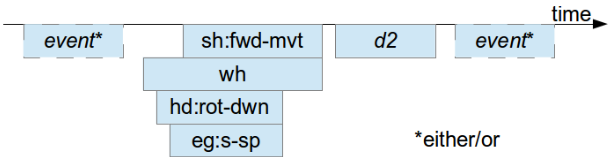

Figure 5: Form diagram for periods between enunciation time and a given time boundary

Figure 6: Snapshot of a form used for periods between enunciation time and a given time boundary

The form of fig. 4 is generally glossed "until", following LSF dictionaries' figure 7. Though, no occurrence actually carried the pictured form: all hand movements were performed sidewards, and every occurrence had a body movement along the manual one[1]. Typically, the change of movement is explained with some form of agreement in location, but in almost no case here could we really come to accept any relevant start and end signing space points for the movements. Besides, to our knowledge, no notice was ever taken of the body movement in such case. It might be argued to be the result of a co-articulation or a phonological process ruling it out as unintentional, but we can only admit that generally speaking very few lateral movements of the hand enrol the body in this way. Incidentally, we note that none of the last two forms use the el:semi-cl feature...

Figure 7: LSF picture dictionary image for "until"

[1]A few examples conforming to fig. 7 were found, but all of them were followed by path end points, not time boundaries. We find that result itself incidentally interesting: are both forms different lexical entries?

What we seem to observe is a propensity to explain manual variations as syntactically driven modification to lexical units, and non-manual additions as non-linguistic, optional or pragmatic. Yet looking at corpus videos with the global approach defended in the previous section, it appears likely that a number of articulators participate in most grammatical functions jointly, and there are non-manual features found inseperable from dictionary units. So, neither denying the existence of lexical units in SL nor that hands play a part therein, we do think there is an unjustified tendency to partition the body into different grammatical roles, the predominant assumption and most deeply rooted idea being a lexical track assigned to the hands.
Beyond the manual channel is of course the non-manual channel, now fairly known to participate in the signed message. Now the point of this paper is to propose that beyond the "manual + non-manual" view, there is yet a visual channel as a whole, with no separation between articulators. Non-manual articulators are articulators in the full sense and should probably not all be grouped under a label and defined by what they are not (manual). Just like ignored minorities claim a right to difference and, once visibility is earned, claim the right to indifference to feel fully included in the system.

2. AZee

AZee is a model to describe and synchronise articulatory forms, built with the philosophy above to synthesise signed productions with a virtual signer, or signing avatar. It comes in the wake of Zebedee, a model proposed a few years prior to this work. Initially made for lexical description, Zebedee:

- allowed writing reusable lexical forms including the invariant forms and the contextual dependencies;
- was based on a synchronisation scheme inspired by Liddell & Johnson's description system of posture–transition alternation (Johnson and Liddell, 2011), developed along the horizontal axis in figure 8;
- made exclusive use of necessary and sufficient articulatory constraints, i.e. no Stokoe-like parametric value was mandatory, only the required articulations were to be specified—vertical axis in the figure.

Especially with the last property above, Zebedee did away with fixed parameters and allowed a flexible articulatory description. However, we have seen that when studying all articulators and all grammatical functions, many features do not perfectly align in body postures but consistently precede or follow, say, a manual movement. Zebedee remained limited in that respect because its focus was still on lexical description, therefore on stabilised, hence mostly time-aligned movements.
To address this problem and gain more expressive power in articulator synchronisation and non-lexical description, the AZee extension was proposed, to:

- enable generic functional rules (whether or not lexical) and their associated forms, including invariant and context-dependent specification;

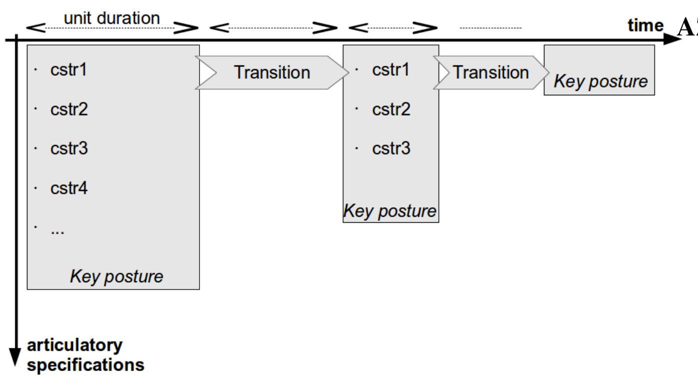

Figure 8: Zebedee

- specify any articulation (whether or not manual) at any time relative to another, for general specification on the time axis.

The basic instruments of the model are a set of native types and a set of typed operators and constants to build expressions normally resulting in XML specifications of animations to synthesise with a software avatar engine. The full set of types is : NUM, BOOL, VECT, POINT, LIST, SIDE, BONE, CSTR, SCORE, AZOP. All are described below.

NUM Numerical values, such as 6.2 or -9.

BOOL Truth values, either TRUE or FALSE.

POINT Points of the signing space. As continued from Zebedee, the signing space in AZee is regarded as a geometric Euclidean space, in which geometric objects can be built as needed and body articulators constrained with.

VECT Vectors of the signing space.

LIST Lists of AZee expressions.

SIDE Left vs. right.

BONE Articulators animated by joint rotations, e.g. the left forearm or the head.

CSTR Constraints that may apply at a point in time, of three main types: bone orientation and placement (forward/inverse kinematics), morphs (for non-skeleton articulators like facial muscles), and eyegaze direction.

SCORE Animation specifications, normally the result of an expression to be used as synthesis input. The only type to cover time, CSTR being articulatory but instantaneous. An XML description excerpt is given in figure 9. It basically specifies a list time-stamped keyframes in a first section, and a list of articulations and morph values to be reached at given keyframes, or held between given keyframes. The basic idea is that any articulator not given a morph value or a joint rotation may be interpolated to reach its next state, or simply take a rest or comfort-selected position.

AZOP Equivalent to functions in functional programming languages. They are to be applied to named argument expressions and result in new expressions. They are most useful to write production rules with non-frozen signed output. For instance, while a shoulder shrug gesture or some non-modifiable sign may be frozen thus described as a SCORE directly, most grammatical rules will be AZOPs with named arguments—such as *duration* in most rules discussed in this paper—and a SCORE output, whose expression depends on the arguments.

```
<Score>
  <KeyFrames>
    <KeyFrame id="1000" time="0"/>
    <KeyFrame id="1001" time="0.0001"/>
    <KeyFrame id="1002" time="1.0001"/>
    <KeyFrame id="1003" time="3.0000999999999998"/>
  </KeyFrames>
  <ScoreSpec>
    <Hold start="1000" end="1001">
      <Rots>
        <Rot joint="left_clavicle" x="-0.198961" y="-
        <Rot joint="right_clavicle" x="0.0075564" y="
        <Rot joint="left_shoulder" x="-0.216117" y="-
        <Rot joint="right_elbow" x="-0.0310633" y="0.
        <Rot joint="right_shoulder" x="0.540892" y="(
        <Rot joint="left_elbow" x="0.00735438" y="-0.
      </Rots>
    </Hold>
    <Hold start="1002" end="1003">
      <Rots>
        <Rot joint="left_clavicle" x="-0.373514" y="-
        <Rot joint="right_clavicle" x="-0.132402" y="
        <Rot joint="left_shoulder" x="0.192286" y="-(
        <Rot joint="right_elbow" x="-0.928085" y="0.
        <Rot joint="right_shoulder" x="0.227843" y="(
        <Rot joint="left_elbow" x="0.772393" y="-0.2(
      </Rots>
    </Hold>
  </ScoreSpec>
</Score>
```

Figure 9: An AZee output of type SCORE

Here is a selection of AZee operators of various argument and result types, which should give an idea of a few things possible with AZee.

`plus`: numerical sum
Type: NUM, NUM → NUM

`scalevect`: vector scaling
Type: NUM, VECT → VECT

`orient`: orientation constraint
Type: str, BONE, str, VECT → CSTR
Articulatory constraint to orient skeleton bones in the signing space. The first argument is either ‘DIR’ or ‘NRM’ depending on whether the bone axis to be oriented is the direction bone (to make it point in a direction) or the normal bone (to lie it in a plane). The second is usually ‘along’ to align the vector in the given vector direction, but ‘//’ is possible to allow opposite direction.

place: placement constraint

Type: site, POINT → CSTR

Articulatory constraint placing a body site at a point in space. The first parameter is a POINT expression, but is not evaluated to 3d coordinates of the point. It must be a body site expression, i.e. one referring to a point on the skin, to be placed at the point given by the second parameter.

morph: morph constraint

Type: str, NUM → CSTR

Articulatory constraint to control non-skeletal articulators such as facial muscles. Morphs have ID names, and can be combined with weights. The first argument is the morph ID to be used; the second is its $[0, 1]$ weight.

key: hold constraints

Type: NUM, CSTR → SCORE

This operation creates the most basic score. A "key(D, C)" expression returns a score of duration D, made of two animation keyframes between which the enclosed constraint specs C will be held. D can be zero, and C can hold any set of constraints: morphs, orientation constraints, placement constraints...

sync: synchronise scores

Type: name, SCORE, list of (name, SCORE, synctype) → SCORE

This operator is the major addition to the Zebedee model, used to synchronise a list of scores. Each score has a name, referred to by the other scores to specify the way they synchronise with the others. A name can be any identifier string; a synctype is a string from the list below, followed by the appropriate boundaries or durations:

- 'start-at', 'end-at': score is untouched and merged starting or ending at a given time position;
- 'start/end', 'start/duration': added score is stretched or compressed to fit the specification;
- 'start/kfalign': score geometry is abandoned and keyframes are aligned with those of the current score...

azop: create an AZee operator

Type: list of (str, AZexpr), AZexpr → AZOP

The result is an azop that can be applied to a context of named argument expressions, which will produce a result typed according to the last AZexpr given. This last expression generally contains references to the argument names, as would any parametrised function in a programming language. Alternatively, the 'nodefault' string can be given if no default expression makes sense; the argument then becomes mandatory when applying the azop.

apply: apply an AZOP to a context

Type: AZOP, list of (str, AZexpr) → returned by azop

The first argument is the azop to be applied. An azop comes with a list of optional or mandatory named arguments, which together form a context for the azop. The return value and type are given by the azop specification. If the azop is a production rule, it will result in a SCORE.

For example, the expression below describes the azop that models the rule sketched in figure 5, with the event signed first. Indentation denotes a parameter under its operator.

```
 1. azop
 2.   'event'  % argument dependency
 3.   'nodefault'
 4.   'until'  % dependency with default
 5.   empty

 6.   sync     % synchronising 6 boxes
 7.     'WH'   %% weak hand box
 8.     key
 9.       1
10.       place
11.         site
12.           'L_KN1'
13.           w
14.           1
15.         [point expression]
16.       [more constraints: hand cfg...]

17.     'EVT'  %% event box
18.     ref
19.       'event'
20.     'end-at'
21.     'WH:0:-.3'

22.     'DATE' %% time boundary box
23.     ref
24.       'until'
25.     'start-at'
26.     'WH:-1:+.3'

27.     'HEAD' %% head drop box
28.     [describe head drop]
29.     'start/end'
30.     'WH:0:+.1'
31.     'WH:-1:-.4'

32.     'GAZE' %% eye gaze box
33.     look
34.       site
35.         'PA'
36.         w
37.     'start/end'
38.     'HEAD:0:+.1'
39.     'HEAD:-1:0'

40.     'SH'   %% strong hand box
41.     [strong hand movement]
42.     'start/end'
43.     'WH:0:+.3'
44.     'WH:-1:0'
```

Lines 2–5 are declarations of the azop's arguments or contextual dependencies, including their names and default expression if absent on azop application, e.g. on l. 5 where 'until' is given a default empty score value. Lines 7, 17, 22, 27, 32 and 40 each names a part of the full signing activity, all to be synchronised by the `sync` operation. The word "box" here is a reference to the rectangles in the illustrations given in figures 3 and 5. Lines 20, 25, 29, 37 and 42 are sync types, i.e. specify the way in which the containing box is to be synchronised with the previous ones. All '*box*:*kf*:*off*' formatted strings are relative time specifications, creating a new keyframe for insertion if none is present at the specified time stamp. In such string, *kf* is the keyframe number of the identified *box*, from which to *off*set the time stamp. The same way values are indexed in Python lists, keyframe numbers are numbered 0 and up from the first to the last, and -1 and down from the last to the first. Line 39 refers to the final keyframe of the score contained in box `HEAD`; line 43 specifies a positive offset of .3 from the beginning of box `WH`.

This azop can be saved under the reference `"Event will last from now until"` and stored as a production rule capable of turning any (*my_event*, *my_date*) pair of scores into a resulting score, combining all boxed features and meaning that *my_event* will last from now until *my_date*. The expression for it is a simple application of the azop with both of its arguments set:

```
apply
  ref
    '"Event will last from now until"'
  'event'
  [my_event score here]
  'until'
  [my_date score here]
```

The interesting and new thing about this model is that the `sync` operation works with any set of scores and any contained articulation specification, except for anatomically impossible constraints. Nothing has enforced us to animate the hands, and no lexical base stream was needed for description. Evaluating this expression produces an XML specification of joint and morph articulations, as presented in figure 9, to be animated directly. Overall, this means we produce animations directly from semantically relevant rule entries and their contextual arguments.

3. Conclusion

This paper has discussed the fact that non-manual articulators were often either overlooked or segregated from manual activity in signing. Firstly, we have not only proposed that they be considered along with manual articulators, but even that all articulators be equal candidates for carrying meaning in SL productions. Secondly, we have made a case against SL articulator role assignment (i.e. projecting grammatical or syntactic functions to specific articulators), and against the assumption that hands would exclusively carry the lexical role. We propose that instead, all articulators be considered together at every moment, and we have shown that with this approach, articulators often seem to behave jointly for the linguistic functions that surfaced.

Then, to describe the observed signed activity with this recommended philosophy, we have presented the AZee model, extension of its ancestor Zebedee. It is capable of describing SL production rules as well as SL productions. That is, by parametrising description elements, AZee can describe generic and context-sensitive rules associating the signed forms to an established SL function, be it lexical or virtually anything else.

One purpose of AZee is to provide a grammar description model covering all SL features, but the aim of our work is ultimately to synthesise the formal descriptions it enables with virtual signers. The first prototype was built and presented recently (Braffort et al., 2013) through a website interface, and will be improved as we go along searching for new production rules.

4. Acknowledgement

We wish to thank WebSourd® for allowing us to study and cite their video material in figures 4 and 6; see `www.websourd.org`.

5. References

C. Baker-Shenk. 1985. The facial behavior of deaf signers: evidence of a complex language. *Am Ann Deaf*, 130(4):297–304.

R. M. Battison. 1978. *Lexical borrowings in American Sign Language*. Linstok Press.

R.-A. A. Bébian. 1825. *Mimographie, ou essai décriture mimique, propre à régulariser le langage des sourds-muets*. L. Colas, Paris.

A. Braffort, M. Filhol, L. Bolot, M. Delorme, C. Verrecchia, and A. Choisier. 2013. Kazoo: A sign language generation platform based on production rules. In *Sign Language Translation and Avatar Technology (SLTAT)*.

C. Cuxac. 2000. *Langue des signes française, les voies de l'iconicité*. Ophrys.

M. Filhol. 2012. Combining two synchronisation methods in a linguistic model to describe sign language. *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication, Springer LNCS/LNAI*, 7206.

M. Filhol. 2013. A rule triggering system for automatic text-to-sign-translation. In *Sign Language translation and avatar technology (SLTAT)*, Chicago, USA.

G. Hickok, U. Bellugi, and E. S. Klima. 1996. The neurobiology of sign language and its implications for the neural basis of language. *Nature*, 381(6584):699–702.

M. Huenerfauth. 2006. *Generating American Sign Language classifier predicates for English-to-ASL machine translation*. Ph.D. thesis, University of Pennsylvania.

Robert E. Johnson and Scott K. Liddell. 2011. A segmental framework for representing signs phonetically. *Sign Language Studies*, 11(3).

B. Moody, D. Hof, and S. Dumartin. 1986. *La Langue des Signes (tome I) – Histoire et grammaire*. Ellipses (reedited: IVT 1997).

W. Stokoe. 1960. Sign language structure: an outline of the visual communication system of the american deaf studies. *Linguistics, occasional papers*, 8.

F. Vergé. 2001. *Le regard en langue des signes française*. Ph.D. thesis, Université Le Mirail, Toulouse, France.