Juxtaposition as a form feature – syntax captured and explained rather than assumed and modelled

Michael Filhol, Mohamed Nassime Hadjadj
CNRS–LIMSI, Université Paris-Saclay
Orsay, France
michael.filhol@limsi.fr, hadjadj@limsi.fr

Abstract

In this article, we report on a study conducted to further the design a formal grammar model (AZee), confronting it to the traditional notion of syntax along the way. The model was initiated to work as an unambiguous linguistic input for signing avatars, accounting for all simultaneous articulators while doing away with the generally assumed and separate levels of lexicon, syntax, etc. Specifically, the work presented here focused on juxtaposition in signed streams (a fundamental feature of syntax), which we propose to consider as a mere form feature, and use it as the starting point of data-driven searches for grammatical rules. The result is a tremendous progress in coverage of LSF grammar, and fairly strong evidence that our initial goal is attainable. We give concrete examples of rules, and a clear illustration of the recursive mechanics of the grammar producing LSF forms, and conclude with theoretical remarks on the AZee paradigm in terms of syntax, word/sign order and the like.

Keywords: Formal grammar, syntax, AZee

1. Production rules

As any language used productively within a community of users, a Sign Language (SL) is a linguistic system, allowing to express and interpret meaning through a set of underlying rules shared by the members, whether consciously or not. For decades now, SL researchers have taken paths looking to identify those rules, together to specify what can be called a Sign Language grammar. More than a few merely parallel schools of thought, the investigated paths are quite numerous and intertwined, forking and joining at milestone concepts. They include descriptive analyses and attempts to build predictive models, some transferring established features of general linguistic description, others rather choosing not to assume anything from other (often written) languages.

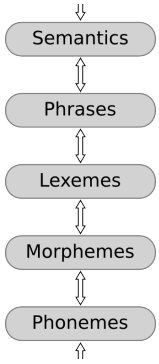

Figure 1: Stacked language construction layer scheme

1.1. AZee: purpose, principle, methodology

The general approach really took off half a century ago with Stokoe (1960). It takes on the stacked layer scheme of figure 1, where every level of language is built from an arrangement of pieces of its lower neighbour. The sensible argument to assume the validity of this scheme for Sign languages is that it was proven robust enough to be considered universal across all studied (written) languages, though it must be admitted that SLs had no fair part in the ones observed as it established.

Describing a language with this scheme implies the possible identification of the layers, each to be formalised with its own dedicated model. In earlier work, we have pointed out the general propensity to explain manual variations as syntactically driven modification to lexical units on the one hand, and to assign special roles to non-manual activity, if it is not simply overlooked, on the other hand.

Yet looking at corpus data with a global approach has revealed that a number of articulators often participate in grammatical functions jointly, whether the articulators are manual or not, and whether the grammatical function is regarded as lexical or not. Also, productive units such as those involving classifiers or complex iconic combinations are still borderline cases whose identification as lexical units versus higher-level productions—or a mix thereof—is a topic of debate, though they represent a significant proportion of the language: up to about half in the annotations of some studies (Garcia et al., 2010).

A few years ago, and initially to bypass this problem and enable synthesis without over- or ill-categorising the language objects, we proposed not to assume the layer stack and fall back on weaker linguistic hypotheses before approaching SL grammar with a formal model, namely:

- language productions are observable *forms* (states and movements of the language's articulators, e.g. "eyelids closed") carrying interpretable *functions* (interpreted purpose or meaning of the production, whether rhetoric, semantic, lexical or unidentified, e.g. "topic change" or "add pejorative judgement on person/object");
- any systematic link between the two is part of what specifies the language, and modelling it yields a rule of the grammatical system.

Such a rule is identified when either:

- an invariant, consistent form is found for many occurrences of an identified function—this raises a *production rule* that can be animated by SL synthesis software;

- a definite function can be interpreted for every occurrence of a certain form criterion—this creates an *interpretation rule*, to be triggered in SL recognition tasks.

Being mostly interested in Sign synthesis, our purpose has been to establish rules of the first kind. The methodology we use to establish those rules is a refining process, consisting in LSF corpus searches for occurrences, alternating between form and function criteria. Starting with a form or function criterion, we list the occurrences satisfying it in the data. If it is a form, give an interpretation (function) for each occurrence; if it is a function, list the form features observed in each case. In either way, group similarities in the new list and use the common features as the criterion for a new search. Alternatively, forms and functions are looked for in the data until one of the above occurs.
A few results have already been published following this approach, as well the descriptive formalism AZee used to describe rules and forms (Filhol et al., 2014). In this paper however, form descriptions will appear as box diagrams, more readable than source code.

1.2. Sequence in delevelled form descriptions

Since the beginning, we have been applying this methodology with the displayed goal of showing how much simultaneity could directly be accounted for when not considering SL streams primarily as sequences of lexical units or glosses. Looking at SL data and describing it with a delevelled approach to grammar and a holistic view of the body, we have shown that simultaneous body movements, head rotations, etc. can often be captured and formally made relevant more easily than, say, attempting to label non-lexical markers on separate levels. Given this intent, we had mostly focused on simultaneous gestures and identified production rules synchronising them.
However, the point was never to mean that sequence was no essential feature. In fact, many rules already described with AZee do produce sequences of signed parts. For example, the form description of our early-identified rule for function "open enumeration of non mutually exclusive items" contains a straight forward sequence of its argument items, only each is decorated with a head movement (see description below).

open-list a.k.a. "etc" (Filhol et al., 2010)

Function:	non-exhaustive list of non mutually exclusive elements
Form:	*items* signed in sequence; forward movement and retraction of the head near the end of each item of the argument list
Example:	*item1* = "scissors"; *item2* = "folding knife"; full interpretation = "scissors, folding knife, and so on"

Incidentally, this rule is established without concern for the lexical or non-lexical status of its arguments. This delevelled parametrisation bears the advantage of allowing to choose them from a lexicon of one-stroke signs as well as to build complex items, which indeed have been observed in such enumerations just as well.
A second benefit of such approach, from which this whole study started, is that it allows to view the ordering of the argument items in the production (the fact that every next item in the argument list is signed *after* the previous) as a particular type of time synchronisation between them, comparable to that synchronising the head movement with each of them. In other words, production of argument forms in sequence is just another form feature available for description.
Our initial point above about simultaneity and the benefit of not reducing grammar to sequential syntax being made, it was time for us now to account for the numerous occurrences of sequences that were not captured by the searches initiated with simultaneous form features.

2. Experiment

In this paper, we propose that time precedence between two pieces of a signed production be regarded as a form feature like any other. Therefore, like one examines head nods or shoulder line rotations in search for their functional motives, one may push the delevelled/holistic approach mentioned above and apply the same methodology to examine sequences of signing chunks and describe their respective functional interpretations. As in the rest of this work, the hypothesis is that consistent appearances of observable forms are the result of intended linguistic functions to be determined, which applied to sequence might eventually account for sign and clause order in a general way.

2.1. Starting with form: juxtaposition

As our methodology goes, the starting point must be a criterion of either form or function, of which corpus occurrences must be listed. In this work, we started with the form criterion of **juxtaposition** of two distinct interpretable pieces (i.e. the beginning of the second occurring after the end of the first), and to describe the functional relationship, if any, that can be interpreted between them from that juxtaposition.
For example, the juxtaposition formed by the chunk meaning "tourist-appealing city" followed by the fingerspelt sequence D-A-H-A-B can be given the function of naming/identifying the former chunk with the latter. By contrast, the sequence of chunks "there were bombs/explosions" and "18 people died, among which several were foreigners" cannot be interpreted that way; the juxtaposition in this case is rather understood as a chronological order of events following that of the production, possibly implying causation in this case.
For each occurrence of that criterion, features of its interpreted function were then given, and sets of common features identified (to serve as the function criterion for the next iteration of the same process, only inverting function and form). After the first iteration, half of the ¡*item1*, *item2*¿ juxtapositions were interpreted as *item2* being a state—in

the most general sense of the term—for *item1*, whether its name, nature, quality or some other complex or detailed chunk of signing that would give information about it.
Following the methodology, this recurrent function feature "state/name/etc. of ... is ..." was made the pivot of the alternation and became a criterion for a new corpus search where forms would be specified in turn. Two patterns emerged involving chin/eyebrow positions and eye blinks, which triggered more iterations of function-to-form and form-to-function searches.

2.2. Results on juxtaposition

The exact statistics for every iteration observed along this corpus study are soon to be published elsewhere, but pulling the thread initiated with the "state" criterion led us to refine it into three stable function-to-form matches, which we summarise below:

category a.k.a. "cat"

Function: *item2* is to be understood as the hyponym of *item1*

Form (fig. 2.2.a): chin and/or eyebrow raise on the beginning of *item1*; minimal transition time (approx. 100 ms) between *item1* and *item2*

Example: *item1* = "country"; *item2* = "Montenegro"; combined interpretation = "Montenegro"

add-info a.k.a. "()"

Function: *item1* is given the additional side information *item2*

Form (fig. 2.2.b): chin and/or eyebrow raise on the beginning of *item2*; minimal transition time

Example: *item1* = "town/city"; *item2* = "the power is tourism"; combined interpretation = "tourist-appealing city"

info-about a.k.a. ":"

Function: *item2* is the point being made about *item1*

Form (fig. 2.2.c): longer transition time (˜340 ms); eye blink at the end of *item2*

Example: *item1* = "power"; *item2* = "tourism"; combined interpretation = "the strength/power is tourism"

Aside from the "state" function leading to the rules above, other functions were interpreted from the occurrences of juxtaposition after the first search, of which we give a few recurrent examples below. However, those need to be refined with a similar process since no further iteration was carried out over them. In other words, while the ones above are consistently observed, the ones below are still hypothetical and need to be studied in more depth by means of more iterations of the methodology. The title functions may merge or be split according to the further observations that will be made.

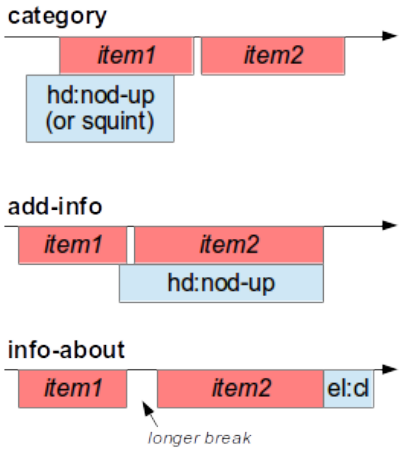

Figure 2: AZee box diagrams for parameterised forms of functions (a) *category*, (b) *add-info*, (c) *info-about*

context a.k.a. "ctxt"

Candidate function: *item1* is the time or space where *item2* is true or taking place

Example: *item1* = "today, May 3"; *item2* = "presidential elections"; full interpretation = "there are presidential elections today, May 3"

NB: The functional distinction was difficult to make with structures where *item1* anchors an entity in signing space, subsequently referenced through their location anchor in *item2*. We therefore hypothesise that these two operations may be covered by the same function.

loc-from-ref a.k.a. "loc"

Candidate function: *item2* is located with respect to *item1* (in time, space or some other projection)

Example: *item1* = placement of Egypt; *item2* = placement of the Red Sea nearby; full interpretation = "the Red Sea, which is near Egypt"

NB: Manual forms appear on the non-dominant hand, in addition to the juxtaposition taking place.

closed-list a.k.a. "and"

Candidate function: exhaustive list of elements with equal part

Example: *item1* = A; *item2* = B; *item3* = C; full interpretation = "A, B and C"

finger-spell a.k.a. "fs"

Function: name a place, person or concept by spelling its written name letter by letter

Example: *item1* = “D”; *item2* = “A”; *item3* = “H”; *item4* = “A”; *item5* = “B”; full interpretation = “Dahab”

NB: This is a rather obvious example of sequence, but it does not come alone. Eye blink before start and eye gaze towards the spelling hand at start seem almost compulsory in our data.

2.3. Elaborating on results

As partly already illustrated in the instances given for add-info and info-about, we find that these rules seem to nest very well, both combining the functions into a correct interpretation and adding up the form features imposed by the rules involved.
For example, the forms we observe for the whole chunk “tourist-appealing city Dahab” are given in figure 3, and are equal to those that would have been combined programmatically from:

1. figure 2.2.c, functionally giving the information “tourism” about the topic “strength”, which builds a chunk meaning “the strength is tourism”;
2. figure 2.2.b, functionally using the chunk above “[its] strength is tourism” as side information to “city”, which builds a new chunk meaning “tourist-appealing city”;
3. figure 2.2.a, functionally making this new chunk a category for fingerspelt chunk identifying the city of Dahab.

Figure 3: Time score of form features for ”tourist-appealing city Dahab”

The combination of rules using recursively built chunks from rules of the same set can easily be represented in a tree diagram, where nodes are rule functions, children are rule arguments and child node order is that of the juxtaposed arguments named *item1*, *item2*... in the descriptions above. The tree corresponding to our example is given in figure 4. This observation about recursion in rule use is consistent with our earlier results on expressions of time sequences and durations, and supports the underlying AZee hypothesis of a recursive system of nestable rules.
Besides, more rules, already published from earlier work, had form descriptions with juxtaposition as the primary form feature, if not the only one documented for the moment. We have already mentioned open-list (§ 1.2.); here are two more examples:

chrono-sequence a.k.a. “seq” (Filhol et al., 2015)

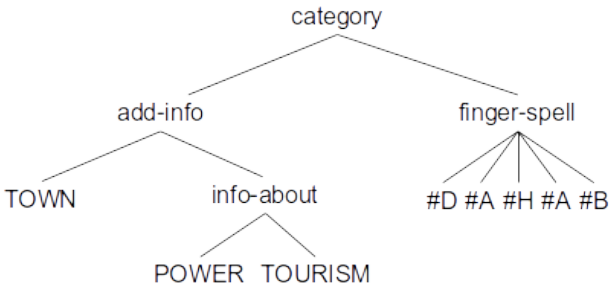

Figure 4: AZee function tree for ”tourist-appealing city Dahab”

Function: *item_i+1* took place after *item_i* chronologically

Form: juxtaposition; more TBD

Example: *item1* = “there were explosions”; *item2* = “people died”; full interpretation = “people died after the explosions”

option-list a.k.a. “either-or” (Filhol et al., 2010)

Function: non-exhaustive list of non-mutually exclusive elements

Form: items juxtaposed; head rest in a new location on each *item* of the argument list

Example: *item1* = “alone”; *item2* = “friends”; *item3* = “family”; full interpretation = “either alone, with friends or with a family”

As the number of established rules grows and by conjoining them into a rule set and generally allowing nesting of one another, we are gradually forming a recursive grammar for LSF, without appealing to preconceived notions like syntax or lexicon.

3. Advantages of the forming grammar

Such delevelled grammar is difficult to situate among the traditional theories, but does provide a few advantages which we underline in this section. They will be useful to discuss the comparison with traditional syntax in the following section.

3.1. Coverage and productivity

Following our methodology, one can at best positively identify function-to-form rules (production rules) and appreciate them for their semantics and assess their coverage. But in view of building a full formal LSF grammar, one can only hope for the existence of a greater system, capturing the whole of the language and of which the identified rules would all be part. Until every corpus discourse becomes describable as a tree of functional nodes taken from an identified set of rules, the question remains of whether such a set exists, and if so what size it might grow into. Is it in reasonable reach at all?
At the beginning of our function-to-form rule search approach and after the first few rules had surfaced, we believed that in addition to those describing the more or less

fixed forms typically listed in dictionaries (the so-called "signs"), hundreds of production rules might be needed given the fine-grain semantic nature the functions we had identified: "open list of non mutually exclusive items", "event separation time exceeding a fortnight", etc.
However, using all identified rules from prior results and from this new study, we have tried to build the trees representing the recursive juxtapositions in 7 videos of 30 seconds each. We have found that out of 220 juxtapositions, only 21 remained unexplained, i.e. about 3 per 30-second discourse. Incidentally, many of them were cases of discourse continuation with hands retracted in a long pause and no semantically or rhetorically loaded function interpretable with consistency. Therefore with only half a dozen extra rules, the new state of our grammar accounts for most occurrences of juxtaposition in our corpus. In other words, it already captures most of what is otherwise called syntax. The exact total number of production rules governing LSF sequence (if any such number exists) may fluctuate according to the various refinements still needed. But if we assume that the refinement searches to come will not break our tentative functions up into big numbers of finer rules ones, our study allows to hypothesise that it should hardly exceed 20 (this would already double the current count), plus dictionary signs.
With this study, our approach to grammar has taken a huge leap forward regarding coverage. At the same time, the resulting rule count ratio has turned from a somewhat alarming figure to an appealing one.

3.2. Expressiveness and precision

We see that the AZee function-to-form rule approach easily accounts for sequences generally considered syntactic. However, the system does not consider sequence—let alone lexical sequence—itself as the primary form of sentence articulation, nor are rules restricted to describe such sequences. It is *one of the many* describable forms, and quite importantly, rules will:

- not only constrain "before"/"after" but include any relevant timing indications;
- contain all other necessary markers directly in their form description.

For example, the add-info rule was observed:

- not only to have the 2nd argument signed "syntactically after" the first argument, but also to constrain the time duration in between (approx. ˜100 ms, i.e. a quick transition);
- to impose a chin lift (or slight squint) starting immediately before and carrying over onto the 2nd argument.

Whereas the first feature is mostly regarded or discarded as "prosody", we observe enough consistency and bond to semantically categorisable functions to keep them in the grammatical descriptions where relevant. There is no reason here to discard transition lengths from the descriptions if they are as systematic as the other synchronised forms. For the same kind of reason, the second feature saves us from justifying grammatical markers on different levels.
This to us makes rules very expressive and brings a lot of precision to the model, for our original goal of computer synthesis but also in linguistic terms. When animating an avatar, if a form feature is consistently observed for a given function in the language data, it is expected to be rendered in the resulting animation, hence it is crucial to know about it and efficient to include it in the form description directly. Linguistically, considering transition durations as just another form allows to parse an input discourse with this information.
A significantly greater precision is thereby achieved when processing sign streams. From the traditional point of view of gloss sequences, utterances like:

TOWN STRENGTH TOURISM

are ambiguous since nothing differentiates between meanings "strong town called [Tourism]" and "tourist-appealing city". With our holistic approach and an AZee grammar, they are directly distinguishable by the better-informed rules, hence the input to parsing process is not considered ambiguous. All so-called "non-manual" or "prosodic" markers will in fact play their equal part in the process directly, instead of being looked for afterwards as a means of resolving the ambiguity, which calls for yet a different system to model.

4. Discussion on "syntax"

Now that the philosophy of our approach has been clarified and a few benefits explained, this section addresses the tricky comparison with traditional syntax, defined as the paradigm governing word order or, for SL, sign sequences.

4.1. An alternative definition for syntax?

Starting from a simple search for function–form mappings, we end up with a set of production rules for every repeated synchronisation of a form feature set, possibly parametrised with arguments. By design, they are not tied to any labels in terms of the traditionally distinct levels of language construction (lexicon, syntax, etc.). However, the linguistic tradition induces a strong intuition of them and a tendency to assign one and one only to every rule. While this may not be needed as such for Sign synthesis from AZee, one might still have interest in such categorisation.
Since the levels were not assumed before building the rules, and as the rules are formally specified with an unambiguous description system, formal criteria can be expressed to characterise the levels formally from AZee. To do so, one can express binary conditions on the rules' form features to group those that intuitively pertain to the same level.
To characterise syntax for example, inspired by the original idea of it governing the order of components, we would suggest that be called syntactic:

> a rule that accepts at least two mandatory arguments, and whose arguments are juxtaposed in the form description.

This definition is interesting in the sense that it does not depend on the notion of lexical unit, which is itself can be problematic to define (see § 1.1.). This raises interesting

prospects regarding known borderline cases such as "classifier predicates" (Cogill-Koez, 2000) or "partly-" or "non-lexical signs" (Johnston and Schembri, 1999). Providing AZee rules for such constructions and checking them against the proposed criteria should feed the discussions on their debated statuses on the basis of formal and purely data-driven arguments.

4.2. Syntax paradox

The last section investigated how AZee encompasses the traditional definitions above, and suggested that the effort would assist clarifying the lines between historical categories. The present section now takes a somewhat reversed point of view, and addresses the question of whether it is relevant at all in the AZee framework in return, specifically in the case of syntax.

A problem when categorising AZee rules as syntactic or non-syntactic on the basis of presence or absence of argument juxtaposition does not acknowledge the fact that no AZee rule exists in the first place if no meaningful interpretation can be made of it. This is quite opposite to the traditionally accepted profile of a syntactic rule. Syntax normally organises argument units according to their respective morphosyntactic categories and regardless of the semantic roles that they may take, and is what fills the gap between lexical units and the semantic relations between them. Contrarily, AZee rules with arguments will necessarily bear some semantic relationship between them directly, otherwise they will simply not exist. By construction, meaningless arrangements of of rules. So firstly, syntax cannot be defined as a category of rules kept clear of semantics by fear that it would be escaping its field. Such criterion would be paradoxical in AZee.

It is only if we accept that a rule might satisfy more than one level-defining criterion like that of section 4.1., that a category of syntactic rules can be identified. But such category does not lie as a necessary level bridging an open gap between two others, so the new question then is what the purpose of it would be. As noted in section 3.2., AZee rules can organise the form features in a variety of ways, including simultaneous features and differences in intervals even between non-simultaneous (juxtaposed) items. So juxtaposition taken on its own as a criterion for rule categorisation does not appear as a specifically salient feature. The relevance of a category of rules only interested in what is sequentially ordered (i.e. syntax) is therefore to be questioned.

In short, we have defined a grammar model without assuming a distinct syntactic organisation of the utterances, then given a criterion to create a syntactic category nonetheless, only by deriving it from the model rather than assuming it to build the model. We concluded that whereas the AZee criterion might have some relevance to clarify the traditional notions when applied to Sign Language, the category itself had little purpose in the AZee paradigm.

5. References

Cogill-Koez, D. (2000). Signed language classifier predicates: Linguistic structures or schematic visual representation? *Sign Language & Linguistics*, 3(2):153–207.

Filhol, M., Braffot, A., and Matthes, S. (2010). Dictasign deliverable d4.2: Sentence descriptions for bsl, dgs, gsl, lsf. DictaSign project, EU-FP7.

Filhol, M., Hadjadj, M. N., and Choisier, A. (2014). Non-manual features: the right to indifference. In *Representation and Processing of Sign Languages: Beyond the manual channel, Language resource and evaluation conference (LREC)*, Reykjavik, Iceland.

Filhol, M., Hadjadj, M. N., and Testu, B. (2015). A rule triggering system for automatic text-to-sign translation. *Universal access in the information society (UAIS), C. Stephanidis (Ed.), special issue on "Recent Advances in Sign Language Translation and Avatar Technology", R. Wolfe, E. Efthimiou, J. Glauert, T. Hanke, J. McDonald, J. Schnepp (guest eds.), Springer.*, 15/4.

Garcia, B., L'Huillier, M.-T., Frémeaux, A., and Makouke, D. (2010). Discursive and morphological processes in the formation and the stabilisation of lexematic units in french sign language. In *Theoretical Issues in Sign Language Research (TISLR)*, UCL, London.

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

Johnston, T. and Ferrara, L. (2012). Lexicalization in signed languages: when is an idiom not an idiom? In *Proceedings the 3rd UK Cognitive Linguistics Conference, vol. 1*, University of Hertfordshire.

Johnston, T. and Schembri, A. (1999). On defining lexeme in a signed language. *Sign Language & Linguistics*, 2(2).

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

Marshall, I. and Sáfár, E. (2004). Sign language generation in an ale hpsg. In *Proceedings of the HPSG04 conference*.

Stokoe, W. (1960). Sign language structure: an outline of the visual communication system of the american deaf. *Studies in linguistics, occasional papers*, 8.