A First Corpus of AZee Discourse Expressions

Camille Challant, Michael Filhol
Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, 91400, Orsay, France
camille.challant@lisn.upsaclay.fr, michael.filhol@cnrs.fr

Abstract

This paper presents a corpus of AZee discourse expressions, i.e. expressions which formally describe Sign Language utterances of any length using the AZee approach and language. The construction of this corpus had two main goals: a first reference corpus for AZee, and a test of its coverage on a significant sample of real-life utterances. We worked on productions from an existing corpus, namely the *40 brèves*, containing an hour of French Sign Language. We wrote the corresponding AZee discourse expressions for the entire video content, i.e. expressions capturing the forms produced by the signers and their associated meaning by combining known production rules, a basic building block for these expressions. These are made available as a version 2 extension of the *40 brèves*. We explain the way in which these expressions can be built, present the resulting corpus and set of production rules used, and perform first measurements on it. We also propose an evaluation of our corpus: for one hour of discourse, AZee allows to describe 94% of it, while ongoing studies are increasing this coverage. This corpus offers a lot of future prospects, for instance concerning synthesis with virtual signers, machine translation or formal grammars for Sign Language.

Keywords: AZee, Sign Language, Formal representation, Corpus

1. Introduction

Formal Sign Language (SL) description is often an essential piece for any kind of processing, especially for synthesis with virtual signers. The most common way to represent SL is based on models developed for spoken languages which have a writing system. They are concerned with the order of constituents in the discourse, thus assuming the existence of different linguistic levels (phonological, lexical, syntactic, etc.) and a sequence order (Napoli and Sutton-Spence, 2014; Padden, 2016; Pfau et al., 2018). This raises several problems, for instance because some structures in SL are difficult to identify as clearly belonging to a specific linguistic level, or because the multi-linearity of SL and a linear sequence are not always compatible. This paper deals with work done on a formal model named *AZee* (Filhol et al., 2014), which avoids most assumptions on language if only coming from spoken languages. The basic principle assumed is that a language creates strong associations between forms and meaning. It allows to take into account many specific features of SL, related to their visual-gestural modality. As AZee began to prove itself for synthesis with an avatar on short but increasingly complex examples (Filhol and McDonald, 2020), we wanted to scale up the test and address SL discourse longer in duration.

The first goal of our study is therefore to test the coverage of AZee, by confronting the model with a large amount of data for the first time. We also want to create an AZee reference corpus which can be used by the scientific community to begin to apprehend AZee.

In this paper, we briefly summarise the AZee formalism, present the corpus we worked to encode and the way in which we built the expressions representing the utterances. We then present the resulting expression set and evaluate it, before ending with a few prospects.

2. AZee

AZee is a formal SL representation approach based on two separate elements. The first is a native functional language capable of describing SL forms to produce, i.e. multi-linear body articulations and their synchronisation or precedence. It defines a set of basic types like string or numerical, but also more SL-specific ones like geometric vectors and points, useful to address signing space. Most notably, values of type score capture timelines of signing activity, synthesisable by an avatar, and type AZop is the functional type, i.e. whose values are functions that can be applied to arguments.

The second element is the notion of "production rule" for a given SL, i.e. a strong association of systematically observable forms (set of articulators and the way they are synchronised or arranged in time) with their interpreted meaning. Production rules can have mandatory or optional arguments, which can be of different types (Hadjadj et al., 2018).

A methodology has been developed to identify production rules in SL corpora. It consists in alternating search criteria of form and meaning until strong pairings establish. For example:

- the form shown in figure 1 associates with the meaning "cold, winter" in French Sign Language (LSF);
- the synchronisation of forms illustrated in fig. 2 associates with the meaning "*info*, given about *topic*".

Such rule therefore surfaces from the study of SL data only. No rule is assumed to exist beforehand.

To encode these form-meaning associations, we use the native type AZop to create a function whose return value is the form to produce, and we assign it a reference. The label is chosen to reflect the interpreted meaning, for example "info-about" for the second association above. As for the first one, a simple ID-Gloss can be used (Johnston, 2010), in this case "*froid*" (cold in English).

Figure 1: A form meaning "cold" or "winter" in LSF (IVT, 1997)

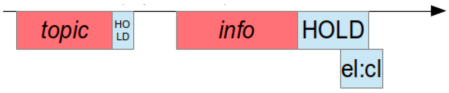

Figure 2: A synchronisation of forms meaning "*info*, given about *topic*"

We call "production set" the set which contains all the production rules found for a SL. Expressions can then be constructed to describe full SL utterances, by combining rules taken from this production set. We refer to these expressions as "AZee discourse expressions". For example, the expression below generates the utterance meaning "it will be cold tomorrow"[1].

```
:info-about
  'topic
  :demain
  'info
  :froid
```

As some production rules in the set allow recursion, we can in principle build AZee expressions of any size, to represent discourses of any length.

3. Corpus selection

The available corpora in LSF are not numerous. Among the recent LSF corpora, we can all the same cite *LS-Colin* (Cuxac et al., 2014), a part of the *Dicta-Sign* corpus (Efthimiou et al., 2012) (LIMSI, 2020) and *les 40-brèves* (Filhol and Tannier, 2014) (LIMSI, 2012).
The last one (downloadable from Ortolang[2]) is an elicited corpus of 40 short news items in written French, which have been translated in LSF by three deaf professional translators each. 120 monologues were thus recorded, for a total duration of one hour.
The journalistic genre is particularly interesting to us, because:

- the productions are monologues facing the camera (no interruptions or feed-back noise between multiple signers);
- it deals with many different topics (non-restricted world);
- it strives to ensure error-free canonical language (there are no disfluencies in the productions).

[1] In French: *demain* means tomorrow, *froid* means cold.

[2] `https://www.ortolang.fr/market/corpora/40-breves`

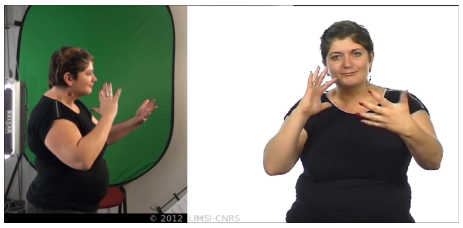

Figure 3: Screenshot of a video from the corpus

From a technical point of view, the videos are encoded in 25 fps and present synchronised front and side views of the signer, as shown in Figure 3. The good quality of the videos enables to observe the movements produced by the signers in detail.
To summarise, this corpus was an appropriate choice for our test of the coverage of LSF by AZee. The next section explains how we went about encoding it.

4. AZeefication

We called "AZeefication" the process of creating an AZee discourse expression for a given SL production (video). The task is to combine production rules to build an expression both capturing the meaning of the utterance and producing all of its observed forms. This section presents a convenient way to achieve this, using the 2L-VF video entry to exemplify each step.
First, we identify non-overlapping interpretable segments in the video. For example, taking 2L-VF in viewing order, one may begin with identifying the segments meaning:

1. "date/moment" (0.00s–2.28s);
2. "cold/winter" (3.08s–3.48s);
3. "when are the sales?" (3.52s–5.20s);
4. "Wednesday" (5.44s–5.76s);
5. "in the whole of France, shown on a map located on the left-hand side of the signing space" (5.80s–7.64s);
6. "on the internet at midnight, located on right-hand side of the signing space" (7.92s–10.64s);
   etc.

These segments each correspond to a sub-expression of the target full discourse expression. The segmenting of the video therefore recursively creates as many simpler AZeefication sub-tasks (developing down each segment), plus one of connecting them into a single expression (building up the discourse expression).
Finding a sub-expression for a segment can be done by identifying its root operation, such as an applied production rule. A trivial case for this is that of an applied rule with no arguments. For example, segments (1), (2) and (4) above are covered (in form and in meaning) by the application of non-recursive rules `moment`, `froid` and `mercredi` respectively. Such segments carry a strong "lexical" feel, for their expressions match what would traditionally be single annotated glossed units.

By contrast, a recognised rule can also force to develop expected arguments, and therefore create more nested expressions to be found recursively. For example, segment (5) cannot be captured by a single zero-argument rule because it contains a body turn and relocation applied to its whole content. The perfect candidate root operation for this form is the application of rule `about-point` (with arguments *pt* and *locsig*). Its purpose in meaning is to relate signed information *locsig* to point *pt* in space, e.g. to specify an anchor in space—which is what is done here with France—or refer to it with new information. Applying this rule at the top of segment (5) develops its two arguments, which must in turn be developed (`*`):

```
:about-point
  'pt
  [point where map is located] *
  'locsig
  [the whole of France on map] *
```

Looking at the video, *pt* is `^Lssp`, i.e. a trivial reference to a point in the left-hand half of the signing space. As for the signed score *locsig* covering everything signed at `^Lssp`, it can neither be captured with a zero-argument rule nor is easy to develop from its root. The solution is to go back to segmenting it as we started. Two parts quite naturally appear: (5a) "entire zone (circular sweep over map)"; and (5b) "France".
The latter (5b) is another example of a trivial application of rule `France`. Segment (5a) separates hands on two simultaneous tracks, the weak one holding the corner of an imaginary map open in a frontal plane while the strong hand sweeps over a wide zone specified as France. The adequate rule for this is `landmark-in-place` (arguments *lm* and *sig*), whose meaning is *sig* deployed around landmark *lm*. Developing the expression accordingly, we create two new arguments still to be developed (`**`), and so on:

```
:landmark-in-place
  'lm
  [corner of map on frontal plane] **
  'sig
  [sweep over map zone for France] **
```

In addition to developing down an expression for each created segment and developed argument, sub-expressions must also be connected into an expression capturing all of its segments, and at the top level the whole discourse. To do this, until all segments are connected, we must find production rules which connect at least two segments in such way that both form and meaning match the video. Good clues for form matching are eye blinks, gaze direction change, intentional head movements, posture holds... We generally proceed by eliminating rules that do not fit, either in form or in meaning, until the most suitable one is found.
For example, segments (1) and (2) above can be connected by the production rule `category` (arguments *cat* and *elt*), whose form is given in fig. 4 and whose meaning is "*elt*, to be understood as an instance of *cat*". This is consistent

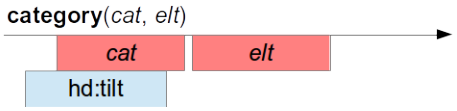

Figure 4: Form synchronisation for rule `category`, arguments *cat* and *elt*

with the form produced around the (1)+(2) segment combination, and with its intended combined meaning "winter": the sign "cold, winter" (segment 2) is polysemic, and the preceding sign "date, moment" (segment 1) disambiguates it. This is a typical use of `category`, likely encouraged by the absence of context, the segment being the very beginning of the utterance.
The AZee expression connecting segments (1) and (2) is therefore:

```
:category
  'cat
  :moment
  'elt
  :froid
```

This reduces the number of segments to connect by one. Similarly:

- segments (5a) and (5b) connect through the same rule `category`, as "France" (5b) is to be understood as a large physical area (5a) in this context;
- segments (4) and (5) connect through the rule `info-about` (see section 2);
- this newly formed combination (4)+(5) connects with segment (6) through rule `each-of` meaning a list of *items*, all given equal focus and truth value.

To summarise, at each step, we either recognise the meaning and form of a segment then have to develop its potential arguments, or recognise segments of the videos then have to find what they are arguments of. We proceed building up and down the expression in this way, until a single expression is built.

However, in some situations, it is not possible to analyse a segment correctly with the available production set. In this case, we decided to briefly describe what the signer meant as if it were a simple application of a rule, and to tag it with a pragma `%E` (for "ellipsis") at the end of the line in the AZee source file. For instance, it is employed when the signers use role shifting or mime something (e.g. "the robber threatens the hostages" in 2A-VF).

5. Results

We applied the AZeefication method above to the 120 videos of the *40 brèves* corpus. The resulting 120 AZee files were added to the original Ortolang repository, each stored in a text file named after the video in the "AZee" directory, thereby creating *version 2* of the corpus[3].

[3]The permanent link to this *40 brèves v2* snapshot is: `https://hdl.handle.net/11403/40-breves/v2`.

In the resulting corpus, we count 11,470 instances of named production rules used in total. The list below summarises the set of those used at least once, restricted to those which have at least one argument of type SCORE (or AZop → SCORE), that is to say that generate recursion. We give the name of each production rule, its arguments and its associated meaning.

- `info-about`(*topic, info*): *info*, which is focused, is given about a *topic*
- `side-info`(*focus, info*): *focus*, with an additional, non-focused *info* given about it
- `category`(*cat, elt*): *elt*, to interpret as an instance of *cat*
- `context`(*ctxt, proc*): *proc* takes place in the context *ctxt*
- `seq-res`(*pre, post*): *post* occurs on the condition that *pre* has happened/finished, or *pre* triggers *post*
- `nicht-sondern`(*nicht, sondern*): not *nicht* but *sondern*
- `each-of`(*items*): list of *items*, each given equal focus
- `all-of`(*items*): set of *items*, with focus on the set as a whole
- `open-list-non-mutex`(*items*): non-exhaustive list of non mutually exclusive *items*
- `mutex-list`(*items*): list of mutually exhaustive *items*
- `fingerspelling`(*letters*): word spelt with *letters* in order in written language
- `sign-supported-sequence`(*units*): *units* concatenated in order in an outside linear system, e.g. dates using day–month–year format, math script reading a formula, sign supported French using words...
- `seq-fleuve`(*events*): list of *events* which follow each other back-to-back in a quick unstoppable succession
- `cam-switch`(*pov*): list of different points of view *pov* on the same event
- `nb-sum`(*nums*): number whose value is the sum of the list *nums*
- `simultaneous`(*sig1, sig2*): *sig1* and *sig2* are true at the same time
- `about-point`(*pt, locsig*): *locsig* about the reference of point *pt*
- `place-object`(*prf, loc*): object denoted by *prf* is placed at location *loc*
- `deploy-shape`(*prf, path*): shape/surface *prf* is deployed along *path*

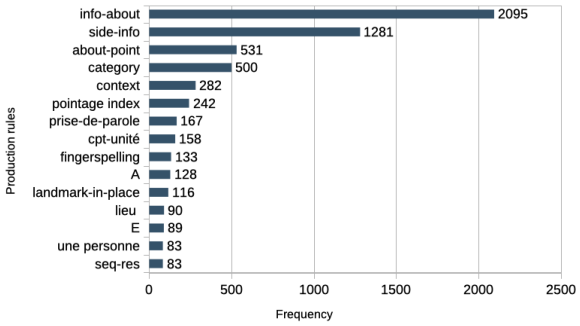

Figure 5: Frequency of the 15 most used production rules in our corpus

- `landmark-in-place`(*lm, sig*): *sig* is constructed around fixed landmark *lm* positioned in space
- `cpt-unité`(*num, elt*): *num* (count) instances of *elt*
- `tens-units`(*tens, units*): the number formed of the two digits *tens* and *units*
- `date-à-date`(*date1, date2*): period of time between dates *date1* and *date2*
- `cpt-années`(*quantity*): age/duration of *quantity* years
- `prise-de-rôle`(*role, sig*): *sig*, in the specified *role*
- `agir-sur`(*ptpatient, actsig*): *ptpatient* is the patient of the action *actsig*
- `pointing-previous-sign`(*ptr, targetsign*): immediately refer back to the just-signed *targetsign* with pointer *ptr*
- `prise-de-parole`(*sig*): recollecting thoughts, taking breath, rhethorical break before *sig* (generates relaxation, hand clasp)
- `double-letter`(*letter*): a doubled *letter* in a spelling sequence
- `insister-sur`(*sig*): insist on, emphasise *sig*
- `pénible`(*sig*): *sig* is painful
- `reverse`(*sig*): reverse meaning of *sig*
- `interruption`(*sig*): *sig* but not quite, or intentionally interrupted
- `regret`(*sig*): *sig* with regret
- `long`(*sig*): *sig* takes time, is long
- `inter-subjectivity`(*sig*): it is generally agreed that *sig*
- `intensity`(*sig*): *sig* is intense
- `soudain`(*sig*): *sig* occurs suddenly

Excluded %E nodes	All	Non-multiplicity
Nodes excluded (count)	334	127
Time excluded (min:s)	3:24	1:03
Total corpus time coverage	94%	98%

Table 1: AZee coverage evaluation

Figure 5 shows the distribution count of the 15 most used production rules. We can see that only a few rules are very frequently used, such as `info-about` and `side-info`, while the others are considerably less frequent, which incidentally recalls Zipf's empirical law known to govern lexical distribution in written languages.

Moreover, it is interesting to note the 15 most frequent rules are quite diverse in number and types of their arguments. For instance, `context` or `seq-res` have two arguments of type score; `pointage-index` has one argument of type point; `lieu` (meaning "place") and `une personne` (person) do not have any mandatory argument; `landmark-in-place` allows to sign several things at the same time, etc. `A` and `E` are rules for letters of the alphabet, used as arguments of the `fingerspelling` rule.

6. Evaluation

One of our main objectives in building this corpus was to test AZee's ability to cover the language, i.e. to bring the pieces not captured by the approach to a minimum. As mentioned in section 4, such pieces were labelled `%E`, acknowledging that further study is needed for them to be properly covered by AZee. By measuring the corresponding video duration of each of these segments, we can measure the amount of corpus coverage lost.

However, many of the `%E` nodes were marked as an application of a non-confirmed rule `multiplicity`. The reason is that they captured a consistent form–meaning association, although with enough variability that we know it should be refined. This work, which mostly consists in applying the methodology mentioned in section 2 for extracting new production rules, is already in progress. From what we have seen, nothing about these nodes seems to break the AZee approach. We are therefore confident these will eventually be cleared, which reduces the number of problematic ellipses. We present both ellipse counts in table 1, respectively including and excluding the `multiplicity` nodes. We can therefore describe most of our corpus with AZee, except for 3 min 24 s, which represents a 94% coverage of the whole corpus duration. Besides, some studies are under way to better understand and analyse the phenomena which are still problematic, in order to further increase this number in the near future.

Moreover, there is still a fine tuning to be done regarding facial expressions. We have already identified such production rules (for example `inter-subjectivity`, `intensity`, `long`...) but others are still known to be missing.

Finally, we keep in mind that we worked on the news genre of discourse specifically, which does not make extensive use of the strongly iconic structures available in SL. The model should therefore be tested with other genres to better evaluate coverage in general.

7. Conclusions and prospects

We described the construction of a corpus of AZee discourse expressions, which extends the existing French–LSF corpus *40-brèves*. We achieved our two main goals, namely to test the coverage of LSF by AZee and to create a reference corpus for AZee. We have formally represented 94% of the LSF discourses with AZee. This result being very promising, this opens up many prospects for future work.

First, we want to study in detail the structures labelled `%E`, to reduce the number of phenomena that cannot yet be described with AZee. Also, we want to continue to test the coverage of AZee on other kinds of corpora, which will enable us to extend the AZee discourse expressions corpus and to further refine the production set. We have already started the AZeefication of another corpus, *Mocap 1* (Benchiheub et al., 2016) (LIMSI and CIAMS, 2020), a part of which consists in describing pictures in LSF and involves many more iconic structures than in the corpus presented here.

An interesting prospect for our work is the test of inter-annotator agreement on an AZeefication task, which was not done with this corpus. This would require comparing AZeefications of the same videos produced by different annotators, and somehow define a metric for it, e.g. edit distance on hierarchical data structures (trees). For the moment, we can only report that when a difficulty in the AZeefication process was encountered, it was always possible, after a discussion between annotators, to find an agreement on the best solution to adopt.

As we explained in section 4, AZee expressions are built by considering both the observable forms and their interpreted meanings. The prospect of automating this process seems complicated at this stage because of the last point. Although, given that a machine can recognise forms more accurately than our human eye, would interpretation of these forms really be necessary in this case?

Moreover, running these AZee discourse expressions through the AZee interpreter allows to compile scores, which contain all the information required by the avatar for reproducing the discourses. Our corpus can therefore be used to test synthesis systems already implementing AZee input (Filhol and McDonald, 2020).

Furthermore, the *40-brèves* is a parallel corpus between written French and LSF and we can consider our AZee discourse expressions corpus as a third parallel entry of this corpus. It is a real positive point because it allows to think about the relevance of AZee in machine translation tasks, as a pivot between French and LSF.

Finally, we will exploit this corpus through statistical studies, in order to measure occurrences or frequencies of production rules, the way they combine, the number and nature of their parameters, etc. It may help us to identify constraints which govern the use of production rules in discourse expressions, from a linguistic point of view. Not unlike property grammars on written language (Blache, 2000), this AZee constraint-based approach could be seen as a new type of formal grammar for SL, based on AZee.

8. Bibliographical References

Benchiheub, M.-E.-F., Berret, B., and Braffort, A. (2016). Collecting and Analysing a Motion-Capture Corpus of French Sign Language. In *7th International Conference on Language Resources and Evaluation - Workshop on the Representation and Processing of Sign Languages (LREC-WRPSL 2016)*, pages 7–12, May 23-28, Portoroz, Slovenia. Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur - UPR 3251 (Limsi), Complexité, Innovation, Activités Motrices et Sportives - EA 4532 (CIAMS) (2020). ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr, v1, https://hdl.handle.net/11403/mocap1/v1.

Blache, P. (2000). Constraints, Linguistic Theories, and Natural Language Processing. In G. Goos, et al., editors, *Natural Language Processing — NLP 2000*, volume 1835, pages 221–232. Springer Berlin Heidelberg, Berlin, Heidelberg.

Cuxac, C., Boutet, D., Dubois, C., and Fiore, S. (2014). Corpus LS-Colin sur plusieurs genres discursifs (Christelle Drecours, Juliette Dalle et Stéphanie Authier). Structures formelles du langage; Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur ; Centre d'analyses et de mathématiques sociales ; Institut de recherche en informatique de Toulouse.

Efthimiou, E., Fotinea, S.-E., Hanke, T., Glauert, J., Bowden, R., Braffort, A., Collet, C., Maragos, P., and Goudenove, F. (2012). DICTA-SIGN - Sign Language Recognition, Generation, and Modelling: A Research Effort with Applications in Deaf Communication. In *International Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (LREC)*, page 5, Valleta, Malte.

Filhol, M. and McDonald, J. (2020). The Synthesis of Complex Shape Deployments in Sign Language. In *Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages*, Marseille, France.

Filhol, M. and Tannier, X. (2014). Construction of a French–Lsf Corpus. In *Building and Using Comparable Corpora, Language Resource and Evaluation Conference (LREC)*, page 4, Reykjavik, Iceland.

Filhol, M., Hadjadj, M., and Choisier, A. (2014). Non-Manual Features: The Right to Indifference. In *International Conference on Language Resources and Evaluation*, Reykjavik, Iceland.

Hadjadj, M., Filhol, M., and Braffort, A. (2018). Modeling French Sign Language: A proposal for a semantically compositional system. In *International Conference on Language Resources and Evaluation*, page 7, Miyazaki, Japan. ELRA.

IVT. (1997). *La Langue Des Signes, Dictionnaire Bilingue LSF-Français*, volume 2. International Visual Theatre, Paris.

Johnston, T. (2010). From Archive to Corpus: Transcription and Annotation in the Creation of Signed Language Corpora. *International Journal of Corpus Linguistics*, 15:106–131.

Napoli, D. J. and Sutton-Spence, R. (2014). Order of the major constituents in sign languages: Implications for all language. *Frontiers in Psychology*, 5:376, May.

Padden, C. A. (2016). *Interaction of Morphology and Syntax in American Sign Language*. Routledge, November.

Pfau, R., Salzmann, M., and Steinbach, M. (2018). The syntax of sign language agreement: Common ingredients, but unusual recipe. *Glossa: a journal of general linguistics*, 3(1):107, October.

9. Language Resource References

LIMSI and CIAMS. (2020). *MOCAP1*. ISLRN 502-958-837-267-9.

LIMSI. (2012). *40 brèves*. ISLRN 988-557-796-786-3.

LIMSI. (2020). *Dicta-Sign-LSF*. v2, ISLRN 442-418-132-318-7.