D4.4: TEXT ANALYSIS MODULES

Revision: v.0.1

Work package	WP4
Task	T4.4
Due date	31/08/2023
Submission date	31/08/2023
Deliverable lead	CNRS
Version	1.0
Authors	Michael Filhol
Reviewers	Özge Mercanoglu Sincan, Fabrizio Nunnari

Abstract	This document presents a justification and specification of the text analysis module intended to perform some level of automatic enhancement of the output of the machine translation to Sign Language, as represented by glosses.
Keywords	Gloss sequence, linguistic information, AZee.

Grant Agreement No.: 101016982
Call: H2020-ICT-2020-2
Topic: ICT-57-2020
Type of action: RIA

Document revision history

Version	Date	Description of change	List of contributor(s)
v0.1	08/2023	Initial version	Michael Filhol
v1.0	15/12/2023	Post-review version	Michael Filhol

DISCLAIMER

COPYRIGHT NOTICE

Project co-funded by the European Commission in the H2020 Programme		
Nature of the deliverable:	**OTHER**	
Dissemination Level		
PU	Public, fully open, e.g. web	✓
CL	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to EASIER project and Commission Services	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

OTHER: Software, technical diagram, etc.

Funded by the Horizon 2020 Framework Programme of the European Union

1 INTRODUCTION TO THE PROBLEM OF SIMULTANEOUS ARTICULATIONS

The point of T4.4 is to leverage formal linguistic knowledge to specify an automatic way of improving the output of the T4.3 machine translation. There are two types of problems or imprecision that can be expected from this output. One is the invalid output compared to what one hoped that the translation model would have been trained to do. For example, when using parallel text and gloss sequences as training data, an incorrect gloss may occur in the output sequence. This is in general considered due the lack of data or its representativeness. The list of possible mistakes of this kind is of course dependant on the data representation and format.

The other comes from the simplifications or approximations of the data representation itself. The issues here do not incriminate the translation module in the pipeline, as they will remain in the output no matter how well-trained the model is. They are rather baked in the choice of representation itself. For example, animating from a gloss sequence used to represent input signed utterances does not properly control features like the time that separates two consecutive signs, the speed at which to produce them, simultaneous body or face articulations, space relocations, eye gaze, etc. It is also limited in handling signed units that cannot consistently be glossed, like size and shape specifications—often termed SASSs in the literature—which make a separate and simultaneous use of the hands and are deployed in continuous space in a relevant manner, without matching a discrete label.

Whereas, these features are known to be important in Sign Language (SL) discourse production, sometimes even essential if they are the only way to distinguish two supporting gloss sequences in meaning. For example, consider the following gloss sequence:

TOWN CASTLE PRETTY ME GO

Depending on the timings and articulations performed by the various body parts in parallel, the same sequence can support several interpretations, sometimes quite different in meaning:

1. I am going to the town with the pretty castle.
2. *{town whose name-sign is CASTLE}* is pretty, and/so I am going there.
3. If [I learn that] the castle in town is pretty, I will go.
4. Between the town and the pretty castle, I [choose to] go to town.

…

Figure 1 shows the differences in produced forms corresponding to each meaning. In each diagram, the horizontal arrow represents the time axis, and the boxes show time intervals during which some part of the signing activity takes place. Specified in the boxes, “el:cl” stands for “eyelids closed”, i.e. an eye blink, “eg:…” specifies a direction for the eye gaze, `Lssp`/`Rssp` are two points on either side of the signing space.

Funded by the Horizon 2020 Framework Programme of the European Union

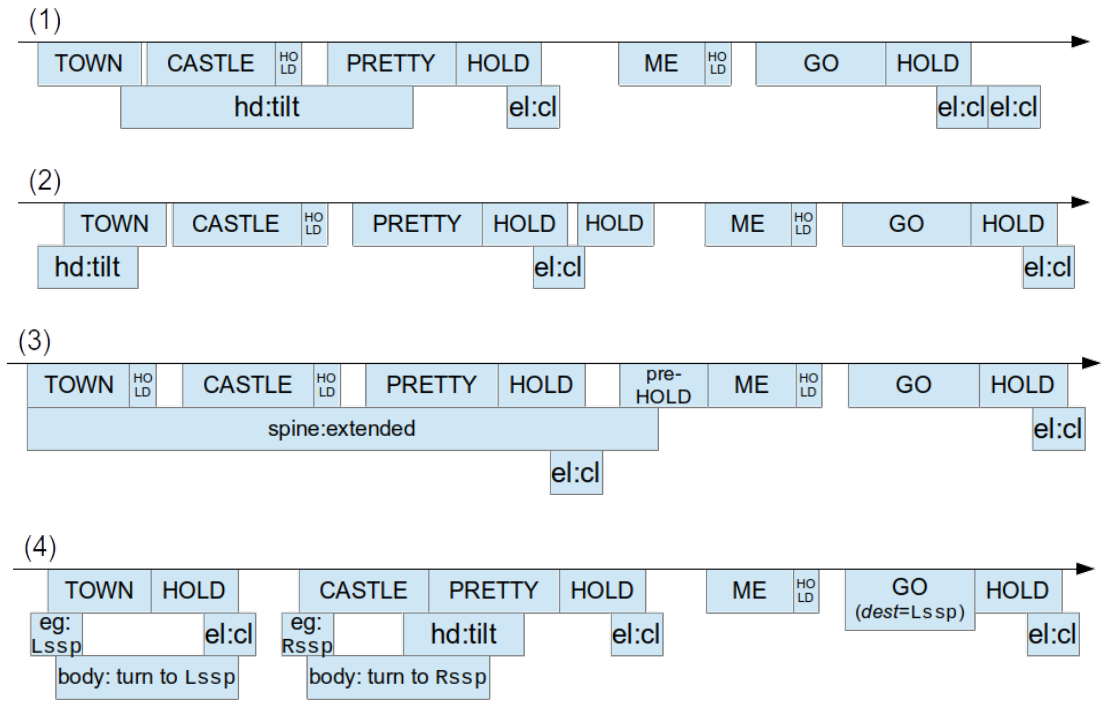

Figure 1: Timelines of 4 distinct signed productions using the same sign sequence

SL streams therefore enclose a lot more than linear arrangements of glossable units alone. Reducing a signed output to a time-ordered list of labels is often convenient because it is a common feature across virtually all of the linguistically annotated data available, and it allows to transfer algorithms developed for linear input like text. But it invariably comes at a cost in expressivity, and is bound to hinder subsequent animations because many parallel features will be lacking. Rendering a concatenation of units (signs) with no further control of the timing between them, or nothing to animate the eye, face, body or mouth in parallel results in a robotic feel and unnatural dynamics that is often not linguitically acceptable (too ambiguous, too unnatural). This task tries to suggest a way of compensating part of this problem by making use of formal linguistic knowledge, drawing from the work done with AZee in the past decade.

2 AZEE

AZee is a formal SL representation approach, aimed at accounting for these subtleties in production, and controlling avatars to synthesise them correctly from semantically informed input. It is based on a native functional language capable of describing SL forms to produce, i.e. multi-linear body articulations and their synchronisation or precedence [1, 2].

To do so, it defines a set of SL-related object types like geometric vectors and points, useful to address signing space. Most notably, values of type SCORE represent timelines of signing activity such as those represented in figure 1, in principle synthesisable by an avatar. Type AZOP is the functional type, i.e. whose values are functions that can be applied to named arguments.

Funded by the Horizon 2020 Framework Programme of the European Union

A second piece of the AZee approach, on a higher level of abstraction and this time for a given SL, is the notion of *production rule*, i.e. a strong association of systematically observable forms (set of articulators and the way they are arranged in time) with their interpreted meaning. Production rules can have mandatory or optional arguments, which can be of different types. Note that 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. Every such rule therefore surfaces from the study of SL data only. No rule is assumed to exist beforehand [5, 8].

For example:

- the form shown in figure 2 associates with the meaning “castle” in French Sign Language (LSF);
- the synchronisation of forms illustrated in figure 3 associates with the meaning “*info*, given about *topic*”, given any two signed pieces *topic* and *info*.

Figure 2: CASTLE in LSF [7]

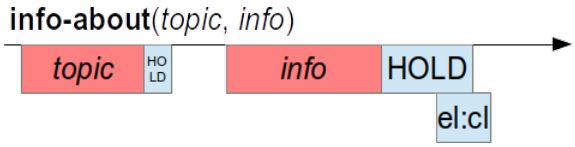

Figure 3: Time arrangement of signed production for rule `info-about`
(red boxes are of variable content and duration, “el:cl” stands for “eyelids closed”, i.e. an eye blink)

To encode these form–meaning associations in AZee, we use the native type AZOP to create a function whose return value is the form to produce (usually a SCORE), and we assign it a reference. The label is chosen to reflect the interpreted meaning, for example “`info-about`” for the second association above. For the first one, a simple ID-Gloss can be used (Johnston, 2010), in this case “`castle`”.

We call *production set* the set which contains all the production rules found for a SL. As some production rules in the set allow nesting of items of the same type (e.g. `info-about` generates a SCORE and itself takes SCORE arguments), we can build expressions of any size, to represent SL utterances of any length. Such expressions, constructed using

Funded by the Horizon 2020 Framework Programme of the European Union

production rules to combine the relevant meanings and produce the appropriate forms, are referred to as *AZee discourse expressions*.

Figure 4: PRETTY in LSF [7]

For example, given a third rule "`pretty`" for the eponymous sign (figure 4) used in our example above, the constructed expression "`info-about(`*`topic`*`=castle(), `*`info`*`=pretty())`" generates the utterance (SCORE) meaning "[the/a] castle is pretty". In native AZee indented style:

```
:info-about
   'topic
   :castle
   'info
   :pretty
```

Building further, consider the "`side-info`" rule which supports the meaning "*focus*, with side, additional or incidental information *info*" (figure 5).

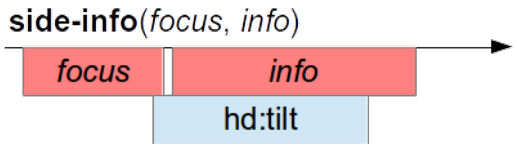

Figure 5: Timeline for rule `side-info`

Funded by the Horizon 2020 Framework Programme of the European Union

Using this rule and the expression in the paragraph above, we can represent the meaning numbered (1) in §1 as follows:

```
:info-about
   'topic
   :side-info     % meaning: town with a pretty castle
      'focus
      :town
      'info
      :info-about % meaning: castle is pretty
         'topic
         :castle
         'info
         :pretty
   'info
   :info-about    % meaning: I go
      'topic
      :me
      'info
      :go
```

The slight head tilt generated by `side-info` (preceding CASTLE and held over onto PRETTY in the final result), the controlled short holds after some signs (none after TOWN, but ~.3 seconds after PRETTY for example), the eye blinks after PRETTY and GO generated by the `info-about` rules... are as many cues that make interpretation (1) the only valid one, as they will contrast with other signals in the other cases—see figure 1. Sub-figure 1.1 is the result of evaluating the expression above.

One sees how this approach captures linguistic knowledge on what to produce on an avatar, handles simultaneity and controls articulator synchronisation in a much finer way than a gloss sequence can approximate. The motivation of this task was to recuperate as much of this knowledge as possible, while still working from gloss sequences, when they are output by the translation system.

3 LEVERAGING AZEE TO IMPROVE GLOSS SEQUENCES

Some distinctions in meaning can only be made through non-glossed cues only, e.g. sentences (1) through (4) in §1. This kind of ambiguity has to be left unresolved in this task, working from the gloss sequence only. What we can hope to compensate is the robotic feel induced by rendering signs one after the other with the avatar.

The general principle with AZee discourse expressions is that nothing is articulated (in form) without being the result of a production rule (with meaning). And sequence being a type of synchronisation, it too should normally be justified by an appropriately selected rule among those available that generate a sequence. And it would then come with some control of the transition time and added body or facial articulations, which would liven up the avatar when doing the synthesis downstream. The question here is that of lifting gloss sequences to a connected AZee expression, in order to approach the naturalness that can come out of it.

Funded by the Horizon 2020 Framework Programme of the European Union

To do so, we propose first to create an artificial AZee rule "`EGG`" (for "EASIER gloss gloss…"), accepting a *units* list as parameter, and whose form is the same (robotic) sequence of *units* as would be concatenated by a sequential animator (figure 6). It is artificial in the sense that it is motivated only by a resulting form—the sequence—, while it is not tied to any identified meaning.

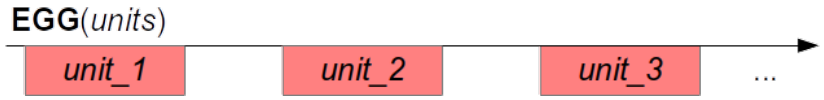

Figure 6: Timeline for dummy `EGG` rule

This way though, any gloss sequence "A B C …" can be wrapped in an `EGG(units=[A, B, C, ...])` expression and thereby be lifted to the AZee-format. For example, the sequence of §1 becomes:

```
:EGG
   'units
   list
      :town
      :castle
      :pretty
      :me
      :go
```

Then, we can identify subsequences in `EGG.units` values which would likely be supported by an AZee composition of properly identified (non-`EGG`) rules. For example, using a lexical resource to look up parts of speech (noun, verb, etc.) from glosses, we can detect the immediate concatenation of a unit glossed with a personal pronoun (like "me") and one glossed with a verb (like "go") at the end of an utterance, and consider that it is likely missing an `info-about` connector. Hence it can be augmented to the more complete expression: `info-about(topic=me(), info=go())`.

Similarly, a noun+adj sequence could be turned into either expression below:

- `info-about(topic=[noun], info=[adj])`, if the adjective terminates—and is probably the point of—the utterance;
- `side-info(focus=[noun], info=[adj])` otherwise, which would apply to the 2nd+3rd element sequence of our example.

Funded by the Horizon 2020 Framework Programme of the European Union

The example EGG expression above, after these transformations, becomes:

```
:EGG
   'units
   list
      :town
      :side-info
         'focus
         :castle
         'info
         :pretty
      :info-about
         'topic
         :me
         'info
         :go
```

By evaluating this new expression, we have taken more control over some of the time separations, and added few body articulations and subtle manual holds, as illustrated in figure 7. This will not salvage any meaning lost in the lack of markings that are not supported by glosses, for example if the production should mean “hesitatingly go” through parallel body dynamics and facial expression on GO. It does however reduce the number of robotic-looking transitions (marked “Ø” in the figure) from 4 to 2, which we propose is a net benefit.

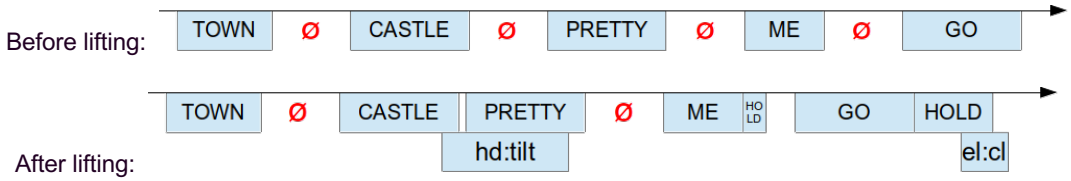

Figure 7: Generated EGG-wrapped scores before (top) and after (b) the lifting process

4 RELATED WORK AND PROSPECTS

The Paula avatar used in the EASIER project has been developing the capability of reading the AZee SCORE format in the recent years, and as demonstrated outside of the project, is now already able to animate many of the specified features [3, 4]. Therefore, although this part of the technology is not planned to be included in the demonstrator, using AZee as a way to lift gloss sequences to a more acceptable rendering seems like a desirable short-term prospect.

Besides, this possibility has motivated the work done to propose an “EASIER notation”, initiated recently by partner UHH in the project (WP6). It can be seen as a way to decorate gloss sequences with new tags and symbols, many chosen in such a way that it becomes possible to build an AZee expression, and in the same way increase the quality of any synthesis from the new input [6].

Funded by the Horizon 2020 Framework Programme of the European Union

REFERENCES

1 C. Challant, M. Filhol (2022). “A First Corpus of AZee Discourse Expressions”, in *Proceedings of the Language Resources and Evaluation Conference*, Marseille, France.

2 M. Filhol, M. Hadjadj, A. Choisier (2014). “Non-manual features: the right to indifference”. *International Conference on Language Resources and Evaluation*, 2014, Reykjavik, Iceland.

3 M. Filhol, J. Mcdonald, R. Wolfe (2017). “Synthesizing Sign Language by connecting linguistically structured descriptions to a multi-track animation system”. *11th International Conference on Universal Access in Human-Computer Interaction (UAHCI) held as Part of HCI International 2017*, Vancouver, Canada.

4 M. Filhol, J. Mcdonald (2020). “The Synthesis of Complex Shape Deployments in Sign Language”, *Proceedings of the 9th workshop on the Representation and Processing of Sign Languages*, Marseille, France.

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

6 T. Hanke, L. König, R. Konrad, M. Kopf, M. Schulder, R. Wolfe (2023). “EASIER Notation – A Proposal for a Gloss-based Scripting Language for Sign Language Generation Based on Lexical Data”. *IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops: Sign Language Translation and Avatar Technology*, Rhodes Island, Greece.

7 IVT (1997). *La langue des signes*. IVT éditions.

8 E. Martinod, C. Danet, M. Filhol (2022). “Two New AZee Production Rules Refining Multiplicity in French Sign Language”. *Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, Marseille, France.

Funded by the Horizon 2020
Framework Programme of the European Union