

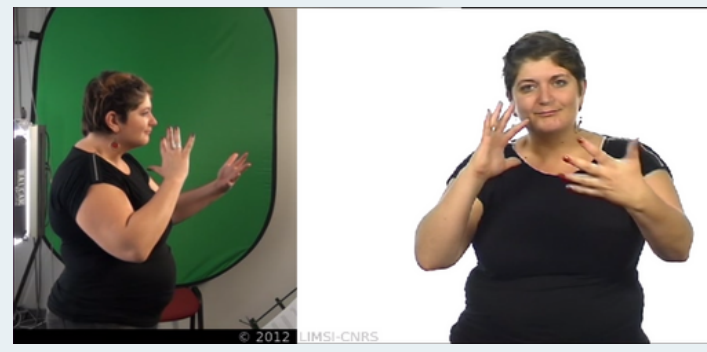
## Context

- Formal description of Sign Language (SL): essential for many tasks of processing
- Traditional models deal with sequence and grammatical categories: poor quality of animation if synthesised
- A model to overcome this problem: AZee

## GOALS:

- test the coverage of AZee on a large amount of data
- create a first reference corpus of AZee discourse expressions

## Corpus selection



"40 brèves" corpus [1]

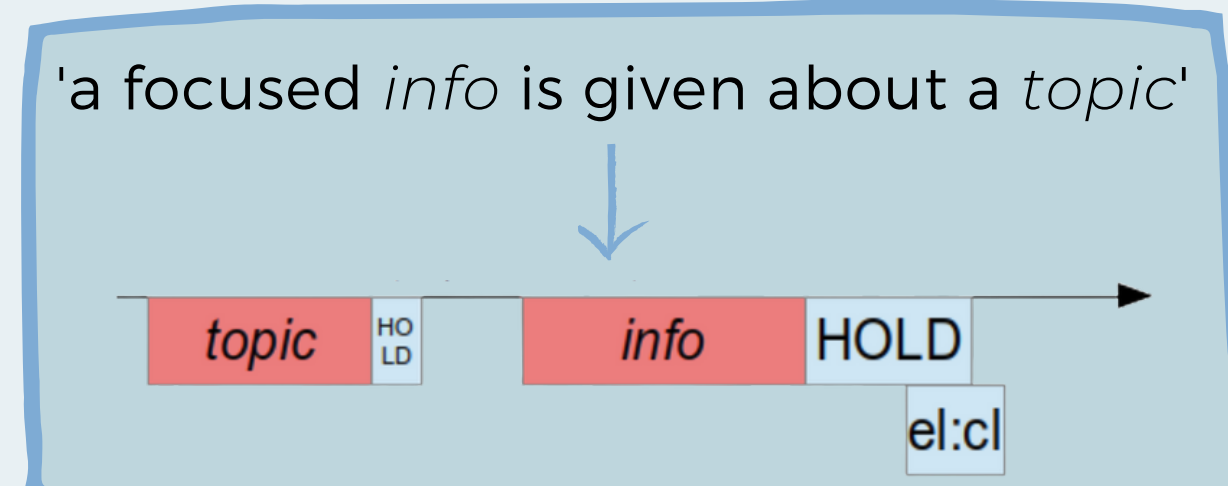
- written French // French SL
- 40 written news items each translated by three deaf professional translators
- 120 SL videos for a total duration of one hour

## AZee [2]

**Production rule:** function representing an interpretable meaning, producing observable forms, possibly parameterised

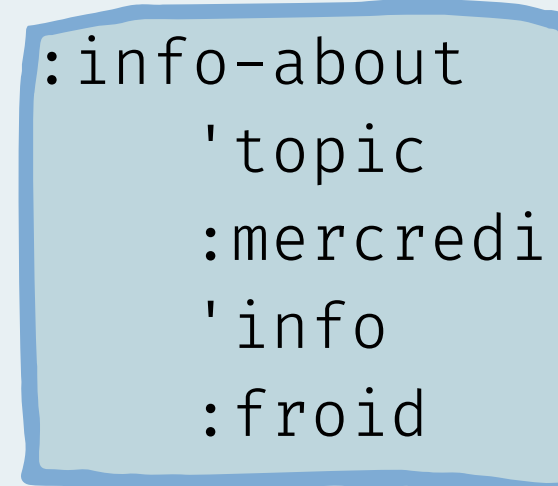


production rule: froid  
resulting score shown as a picture



production rule: info-about  
resulting score shown as a timeline

**AZee discourse expression:** combination of rules taken from the production set, representing a full SL utterance (captures both meaning and form)



> AZee discourse expression meaning  
"it will be cold on Wednesday"

**Production set:** set which contains all the production rules found for a given SL

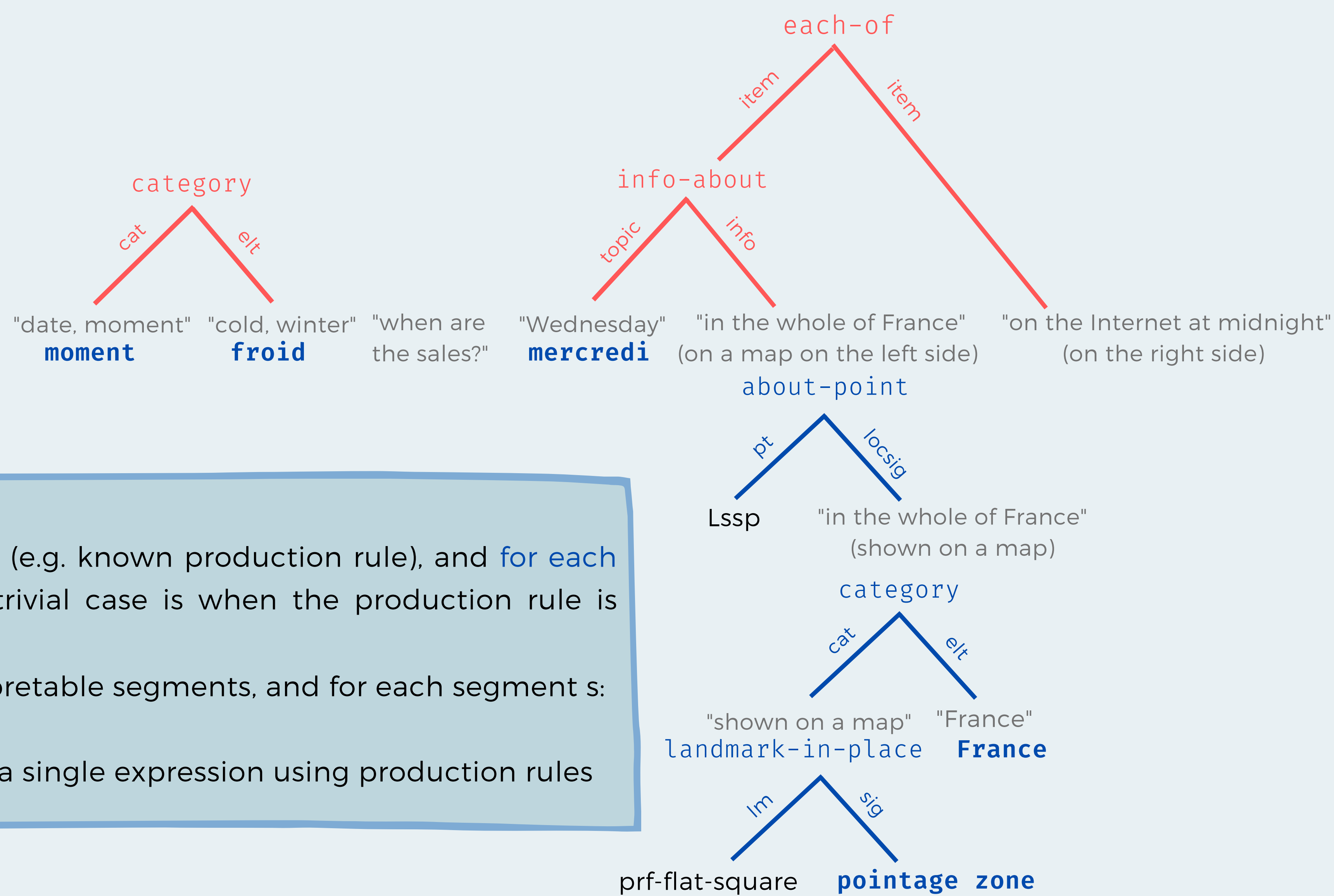
- for example:
- category: elt, to interpret as an instance of cat
  - pointage zone: pointing a target point with a flat hand
  - mercredi: Wednesday
  - ...

**AZeeification:** search for the best AZee discourse expression for a given SL segment

## AZeeification

- Example (2L-VF shortened):

"Winter sales start Wednesday in the whole of France and at midnight on the Internet"



- Process summary:

AZeeify(X):

- recognise a root node for X (e.g. known production rule), and for each argument x: AZeeify(x'). A trivial case is when the production rule is applied to no arguments.
- otherwise, split X into interpretable segments, and for each segment s:
  - AZeeify(s)
  - connect the results into a single expression using production rules

"Wednesday, in the whole of France"

```

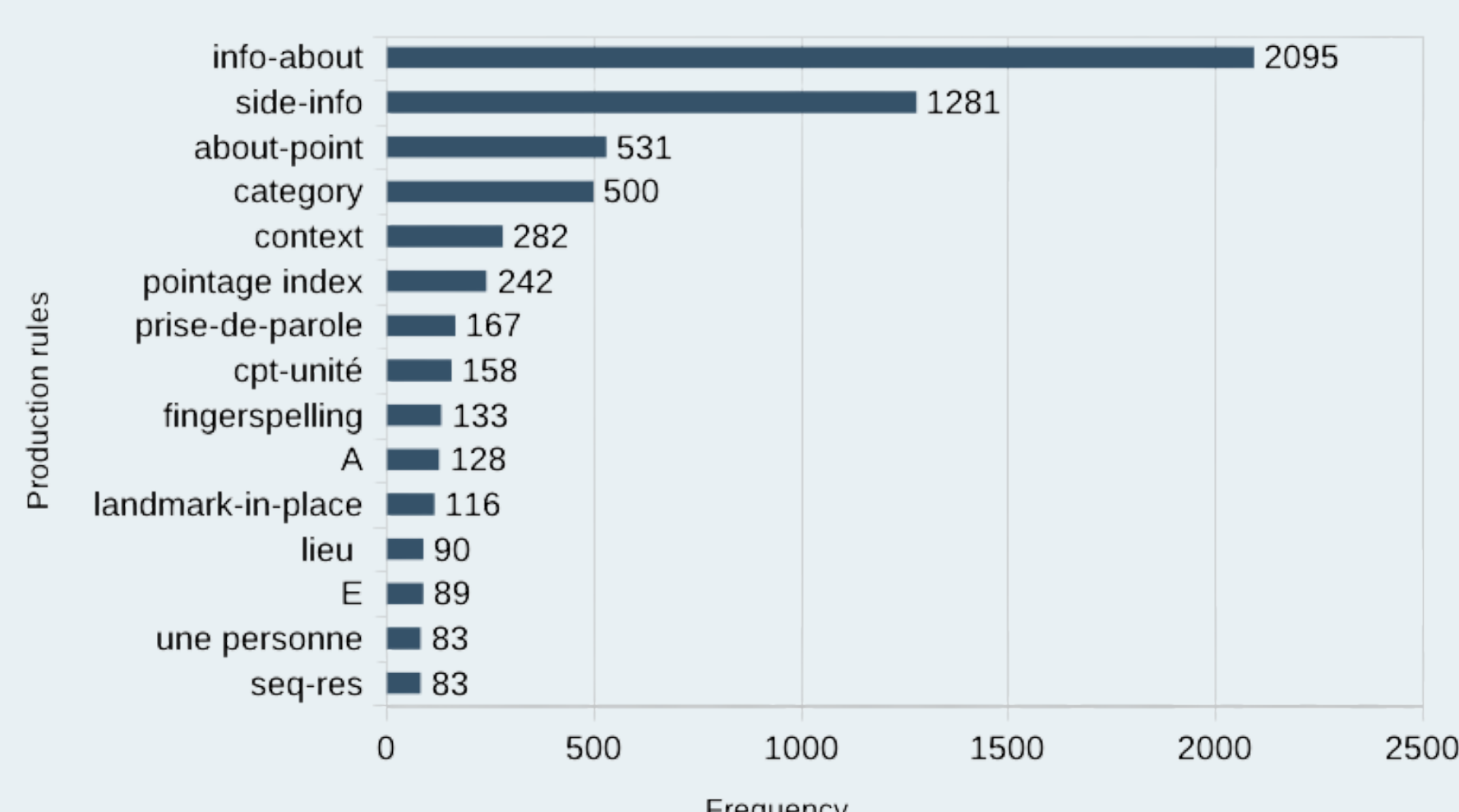
:info-about
'topic
:mercredi
'info
:about-point
'pt
^Lssp
'locsig
:category
'cat
:landmark-in-place
'lm
:prf-flat-square
'loc
^Lssp
'hand
w
'nrmsurface
fwd
'dirsurface
up
'sig
:pointage zone
'target
^Lssp
'size
large
'nrmsurface
fwd
'elt
:France
    
```

## Results & evaluation

- 120 AZee discourse expressions created (1h of SL), available on Ortolang
- 11,470 named production rules used
- Frequency of the 15 most used production rules:



All the segments impossible to describe with AZee at this stage have been identified with %E (for "ellipsis")



	%E	%E without multiplicity
occurrences in the corpus	334	127
time (min:s)	3:24	1:03
AZee corpus coverage (%)	94	98

multiplicity is a meaning with no stable form (study in progress)

94% coverage

## Prospects

SL synthesis with virtual signers, machine translation, formal grammar based on AZee...

## References

- M. Filhol, X. Tannier (2014), *Construction of a French-LSF corpus*. Building and Using Comparable Corpora, Language resource and evaluation conference (LREC), Iceland. LIMSI (2020). 40 brèves [Corpus]. ORTOLANG - www.ortolang.fr, v1, https://hdl.handle.net/11403/40-breves/v1.
- M. Filhol, M. N. Hadjadj, A. Choisier (2014). *Non-Manual Features: The Right to Indifference*. In International Conference on Language Resources and Evaluation, Reykjavik, Iceland.
- International Visual Theatre (1997). *La langue des signes. Dictionnaire bilingue LSF-Français*. Tome 2. IVT, Paris.