

Rosetta-LSF: an Aligned Corpus of French Sign Language and French for Text-to-Sign Translation

Élise Bertin-Lemée¹, Annelies Braffort², Camille Challant², Claire Danet², Boris Dauriac³, Michael Filhol², Emmanuella Martinod², Jérémie Segouat⁴

1



2



3

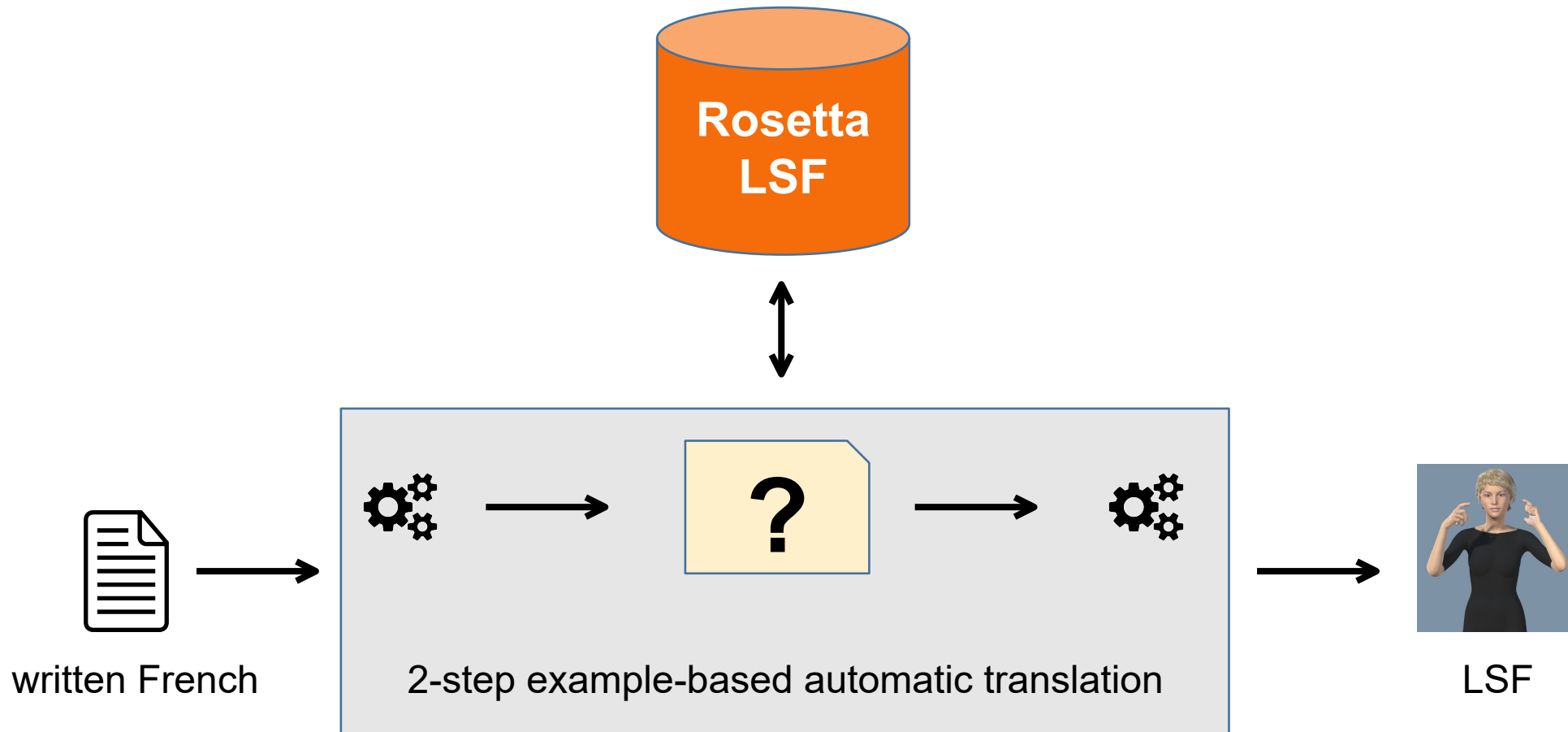


4



<https://rosettaccess.fr/>

Aim: A bilingual aligned corpus to test automatic translation using example-based approach



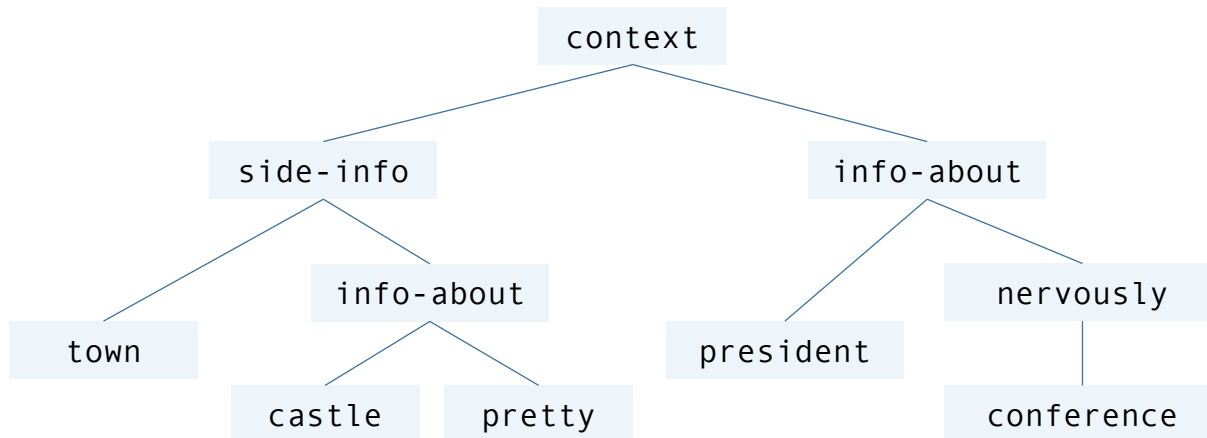
Intermediate representation for text-to-sign translation

- In general : discourse represented as sequence of glosses (sign labels)
- AZee: hierarchical representation based on form—meaning associations

In the town with the pretty castle, the president gave a nervous speech.

TOWN CASTLE PRETTY PRESIDENT CONFERENCE

Meaningful information missing
(facial expr. for “nervously”)



```

:context
'ctxt
:side-info
'focus
:town
'info
:info-about
'topic
:castle
'info
:pretty
'proc
:info-about
'topic
:president
'info
:nervously
'sig
:conference
  
```



Generates determined forms
and their synchronisation,
whether simultaneous or in
sequence

Corpora constitution: signer and setup

- LSF signer
 - a single signer to have consistent dynamics in the mocap data
 - experienced in producing LSF content for the media
- Technical set-up to record LSF mocap data



Corpus elicitation & Content

- 4 tasks including LSF content
 - Parallel content : news titles in French and LSF
 - Description of photos
 - Reproduction of video clips
 - Isolated signs
- Recorded mocap data re-rendered on avatar



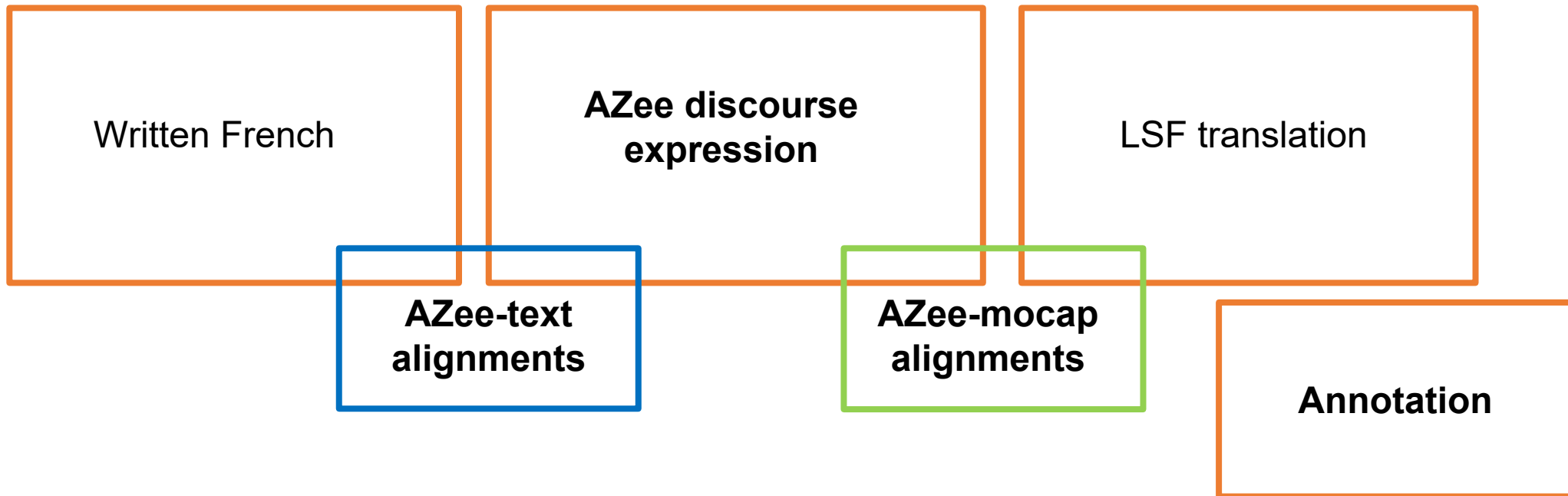
...

Vaucluse : une commune se bat pour conserver son distributeur
 Papillomavirus : les médecins souhaitent une meilleure vaccination
 Rasmus Michau : « Les tâches très répétitives vont être supprimées »
 Affaire Sophie Le Tan : l'ADN de la jeune femme retrouvé sur une scie
 Le propriétaire de la boulangerie a décidé de recycler des invendus .
 Dans un institut médico-associatif qui accueille les enfants autistes .
 Emplois : Carrefour souhaite supprimer 1 200 postes dans ses hypermarchés

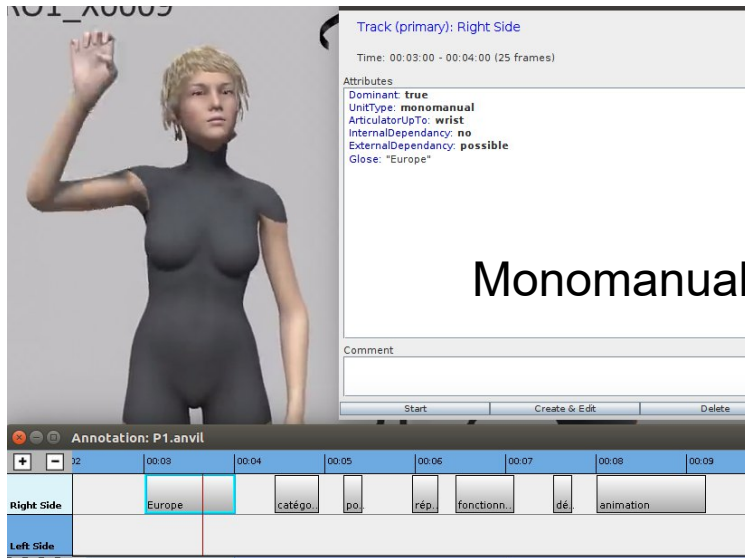
...

Extension of the news title entries for translation

- Extend the news title entries for text-to-sign translation
- Create parallel aligned entries



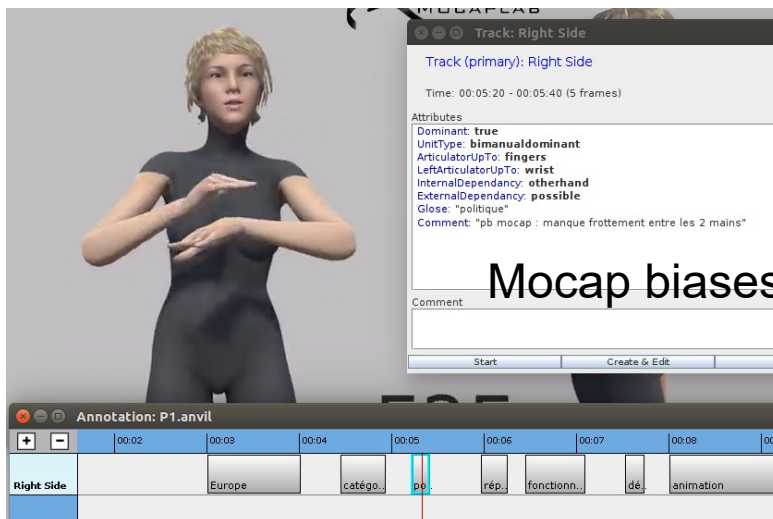
Extension: Annotation



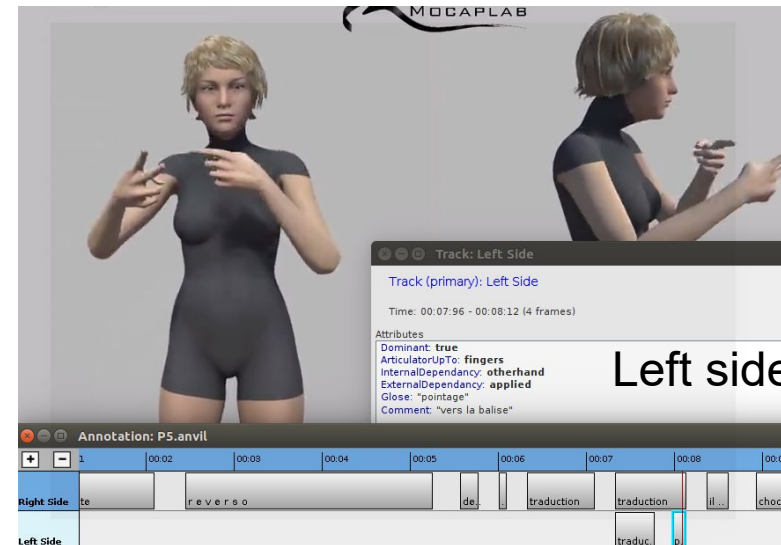
Monomaneuver



Bimanual relocated



Mocap biases



Left side

Extension: AZee discourse expression

- **194 “AZeefications”**

```
:info-about %F ont vendu leur vaisselle %t 1739-1967
  'topic
  :là
  'info
  :info-about
    'topic
    :all-of %F vaisselle %t 1767-1851
      'items
      list
        :assiette
        :assiette
      'info
      :multiplicity %F vendu %t 1855-1967
        'elt
        :vendre
```

→ the basis for an intermediate representation between the French text and the generation of LSF animations

Extension: AZee-text alignments

AZee expression nodes ~ a segment of written French

%F in AZeefications (n = 1812)

- 3 principles:
 - Uniqueness and Maximisation (*a)
 - Objectivity (*b)

```
:category %F Paris  
'cat  
:ville  
'elt  
:Paris %F Paris
```

(*a)

```
:category %F Paris  
'cat  
:ville %F ville  
'elt  
:Paris
```

(*b)

```
:category %F Paris  
'cat  
:ville  
'elt  
:Paris
```

(c)

Extension: AZee-Mocap alignments

AZee expression nodes ~ timed video segments

%t in AZeefications

```
:info-about %F ont vendu leur vaisselle %t 1739-1967
'topic
:là
'info
:info-about
'topic
:all-of %F vaisselle %t 1767-1851
'items
list
:assiette
:assiette
'info
:multiplicity %F vendu %t 1855-1967
'elt
:vendre
```

AZeefication



Corresponding video segment

Example

News title

De grands chefs
Top chefs

AZee expression

```
:category %F de grands chefs %t 1587-1727
'cat
:side-info
'focus
:multiplicity
'elt
:une personne
'info
:pointage zone %l plan horizontal
'elt
:chef cuisinier
```

LSF translation



Annotation

UnitT	Art	IntDep	ExtDep	IdGlose
biequ	wrist	otherh	notcanonical	personnes
mono	wrist	fingers	notcanonical	pointage_circ
biequ	wrist	head	notapplicable	chef_cuisinier

Conclusion

- **Rosetta-LSF** corpus, available on Ortolang



194 news titles - with annotations and AZee



28 picture descriptions



24 LSF video repetitions



~1200 isolated signs

<https://www.ortolang.fr/market/corpora/rosetta-lsf>

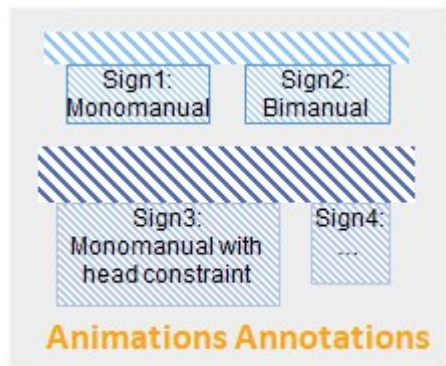
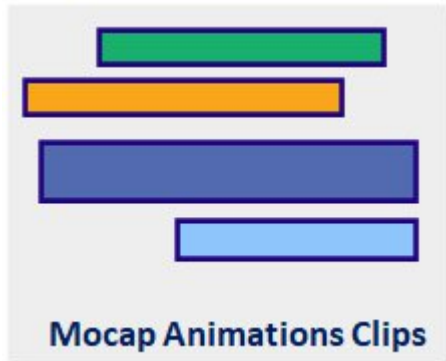
- **Further studies**

- generation
- translation

```

fallback
  category %t
  'cat
  info-about
  ...
  'elt
  info-about
  ...
  category %t
  'cat
  side-info
  'focus
  multiplicity
  'elt
  une personne
  'info
  zone
  'elt
  chef cuisinier
  ...
  info-about %t
  'topic
  exemple
  'info
  info-about
  'topic
  aussi
  'info
  info-about
  'topic
  side-info %t
  'focus
  category %t
  'cat
  ville
  'elt
  fingerspelling
  ...
  'info
  banlieue %t
  'info
  ...

```



Recorded corpus :



194 news titles



28 picture descriptions



24 LSF video repetitions



~1200 isolated signs



ROSETTA project description

