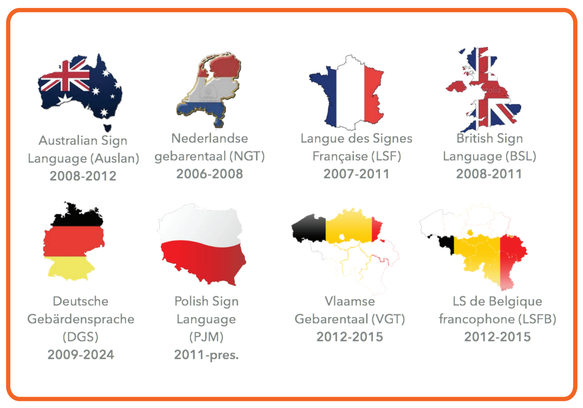
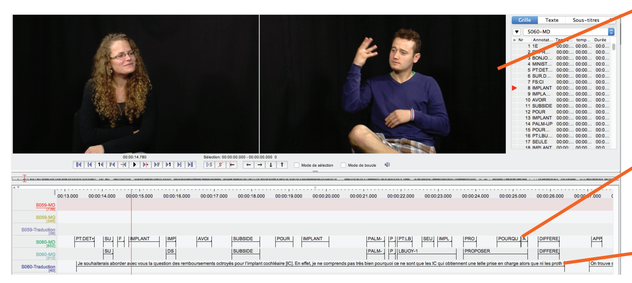
Modelling a Parallel Corpus of French and French Belgian Sign Language (LSFB)

Laurence Meurant[1], Maxime Gobert[2], Anthony Cleve[2]
[1]F.R.S.-FNRS & University of Namur, LSFB-Lab [2]University of Namur, PReCISE research center
laurence.meurant@unamur.be maxime.gobert@unamur.be anthony.cleve@unamur.be



01 Sign Language corpora

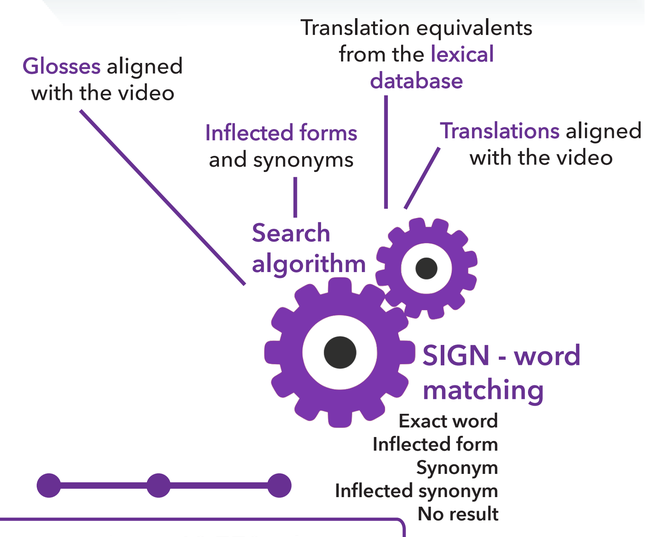


- Videotaped signed data
- Written glosses of the signs (manual task)
- Translation of the data into written language

02 Parallel corpora?

- Sign
- Gloss: IMPLANT
- Translation: des remboursements octroyés pour l'implant cochléaire

03 Alignment



Current results on **18,770 tokens**:

- **6,592**: no result found
- **3,211**: one gloss matching different words (OneToMany)
- **8,211**: several glosses matching the same word (ManyToOne)
  - **5,800**: without the 2-handed signs
- **9,668**: exact match

UNEXPECTED MEANINGS? TO BE REVIEWED REPETITIONS? SIGN-WORD STABLE PAIRS?

04 Meanings in context

AUSSI aussi, même, comme, **aussi que, autant, comme si, égal, également, en plus, même chose, parce que, tout aussi, tout comme, vu que, en quelque sorte, plus**

Additional meanings that enrich the lexical database

www.corpus-lsfb.be

Videos	150 hours x 4 synch. shots
Annotations	12 hours, i.e. 104,000 tokens
Lexical database	2,500 types
Translations FR	3 hours, i.e. 2,400 sentences

Open access corpus of French Belgian Sign Language (LSFB)

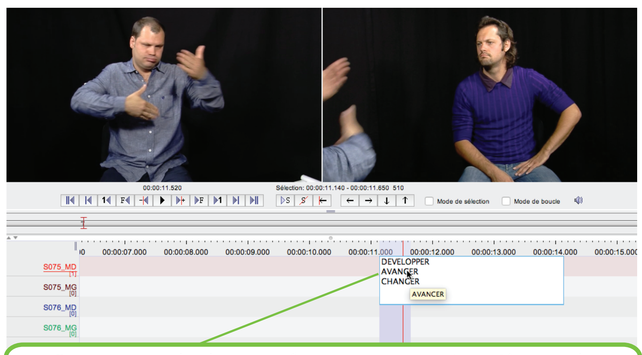
05 Searchable concordancer

Fictitious examples inspired by the Linguee interface

Bank of examples: lexical and non-lexical translation equivalents

06 Assisted annotation

Suggestions to the annotator while annotating



Suggested glosses for one meaning (from the enriched list of meanings in context)

+ Suggestions based on the text of the previously translated video?

UNIVERSITÉ DE NAMUR fnrs LA LIBERTÉ DE CHERCHER

This research is funded by the University of Namur, within the framework of its Research Incentive Programme.

B. Altenberg et al., editors. (2002). Lexis in contrast: corpus-based approaches. John Benjamins Publishing.
Bourdaillet, J., Huet, S., Langlais, P., and Lapalme, G. (2010). TransSearch: from a bilingual concordancer to a translation finder. Machine Translation, 24:241–271.
Deville, G., Dumortier, L., Meurisse, J.-R., and Miceli, M. (2013). Ressources lexicales: Contenu, construction, utilisation, évaluation. In N. Gala et al., editors, Ressources lexicales pour l'aide `a l'apprentissage des langues, volume 30, pages 291–312. John Benjamins.
Filhol, M. and Tannier, X. (2014). Construction of a french-lsf corpus, building and using comparable corpora. In Proceedings of LREC 2014.
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.
Linguee. (2015). http://www.linguee.com.
Meurant, L. (2015). Corpus LSFB. First digital open access corpus of movies and annotations of French Belgian Sign Language (LSFB). Laboratoire de langue des signes de Belgique francophone (LSFB-Lab). FRS-F.N.R.S et Université de Namur.