GeoQuery-LSFB: A French Belgian Sign Language Corpus with Procedural Semantic Annotations

Liesbet De Vos[1], Laurence Meurant[1], Paul Van Eecke[2], Katrien Beuls[1]
[1]Université de Namur, [2]Vrije Universiteit Brussel contact: liesbet.devos@unamur.be

1 INTRODUCTION

- Procedural Semantic Representations
  - describe the meaning of natural language expressions in terms of primitive operations that can be evaluated against images, databases, knowledge graphs or other external sources;
  - are valuable for NLP/NLU tasks (e.g., visual question answering, visual dialogue, database querying, and robot instruction).
- In contrast to spoken languages, no resources with procedural semantic representations exist for signed languages.

Research aim: extend multilingual GeoQuery towards French-Belgian Sign Language (LSFB).

2 MULTILINGUAL GEOQUERY

- 250 questions about US geography in 9 languages: Spanish, Turkish, Japanese, Chinese, Geek, German, Thai, Indonesian, Farsi, Swedish
- Each question is linked to a procedural semantic representation that can be evaluated on a database, retrieving the answer to the question

Natural language Expression

"What is the area of the largest state ?"

Procedural semantic representation

answer(area(largest(state(all))))

Answer

591000.0

3 DATASET CREATION

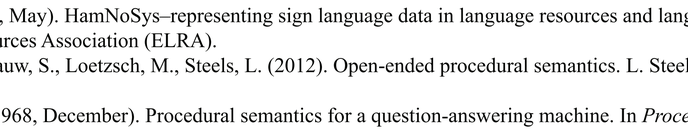
1 Identify templates of meaning representations in GeoQuery

e.g. answer(A,(size(B,A),const(B,stateid(texas))))
answer(A,(size(B,A),const(B,stateid(alaska))))
→ answer(A,(size(B,A),const(B,stateid(<state>))))

2 Translate first question of each group from English into French

e.g. en: How big is Texas ?
→ fr: Quelle est la taille du Texas ?

3 Translate French Questions into LSFB



4 Transcribe and annotate recorded forms

Time

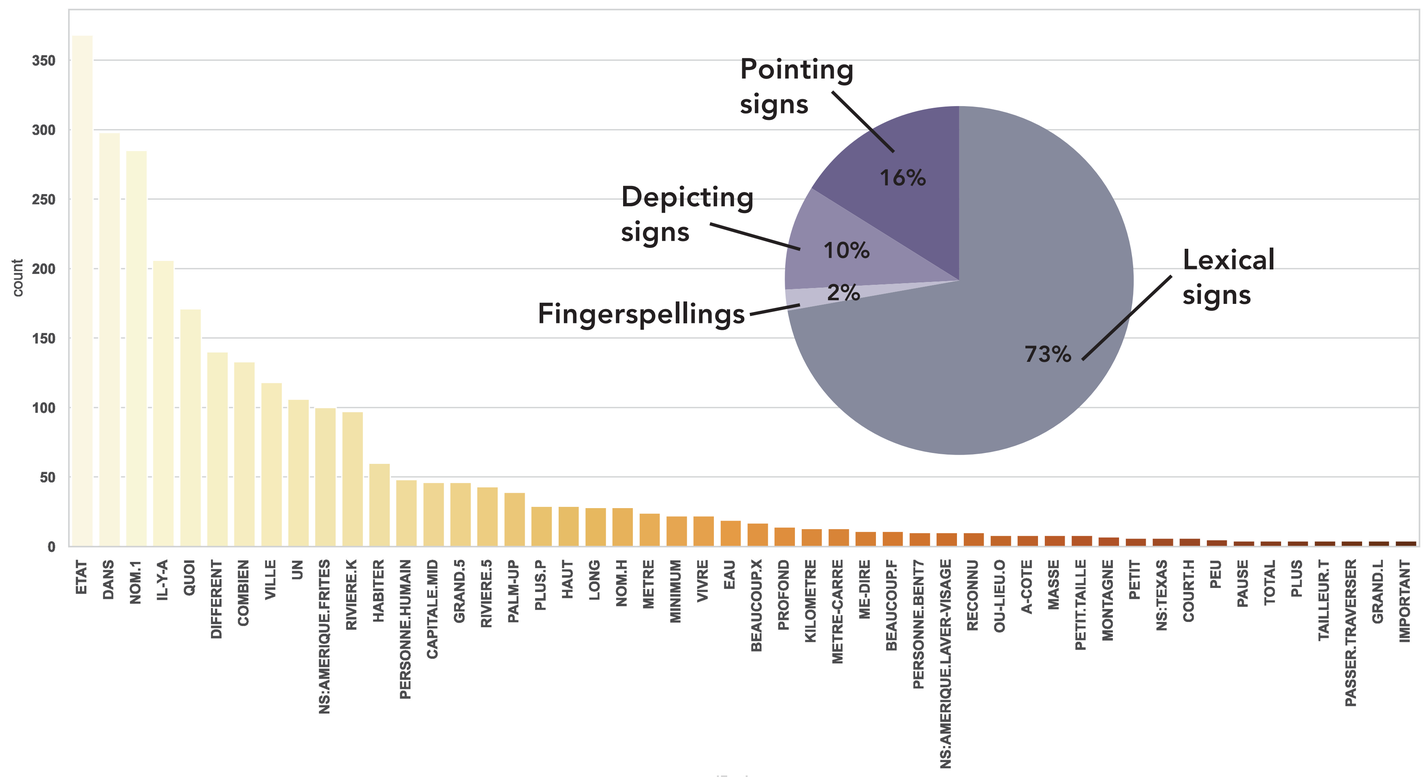
Right hand	PT	NS:TEXAS	PT	SIZE	HOW-MANY
Left hand	DS			SIZE	

5 Extend dataset based on transcribed recordings of templates

e.g. replace NS:TEXAS with NS:ALASKA in the transcribed form

4 DATA DESCRIPTION

- 124 video recordings and ELAN annotation files
- JSON and XML corpus files containing:
  - expressions in all of multilingual GeoQuery's languages
  - translations of each expression into LSFB and French
  - Procedual semantic representations for each expression in three different formats



5 EXAMPLE EXPRESSION

RH ID-gloss	IN	FS:ALASKA		THERE-IS	ONE	CITY
RH HamNoSys						
LH ID-gloss	IN					CITY
LH HamNoSys						

RH ID-gloss	MOST	BIG	HOW-MANY	PT:DET/LOC(1)	SQUARE-METERS	SIZE
RH HamNoSys						
LH ID-gloss		BIG				SIZE
LH HamNoSys						

6 CONCLUSION

- GeoQuery-LSFB combines machine-readable transcriptions of signed utterances with procedural semantic representations.
- Unique resource for the comparative study of spoken and signed languages, both from a theoretical and an application-oriented perspective.
- possible uses:
  - language processing tasks such as semantic parsing, database information retrieval, sign language recognition and machine translation
  - the onomasiological study of a signed language along a typological variety of spoken languages

More information and corpus download:

https://doi.org/10.5281/zenodo.19220833

Further reading:

- Beuls, K., Van Eecke, P. (forthcoming). Humans Learn Language from Situated Communicative Interactions. What about Machines? *Computational Linguistics*.
- De Vos, L., Nevens, J., Van Eecke, P., & Beuls, K. (2024). Construction grammar and procedural semantics for human-interpretable grounded language processing. *Linguistics Vanguard*, (0).
- Hanke, T. (2004, May). HamNoSys–representing sign language data in language resources and language processing contexts. In *sign-lang@ LREC 2004* (pp. 1-6). European Language Resources Association (ELRA).
- Spranger, M., Pauw, S., Loetzsch, M., Steels, L. (2012). Open-ended procedural semantics. L. Steels & M. Hild (Eds.), *Language grounding in robots* (pp. 153–172). New York, NY, USA: Springer.
- Woods, W. A. (1968, December). Procedural semantics for a question-answering machine. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I* (pp. 457-471)