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

Anonymous ACL submission

Abstract

Procedural semantic representations describe the meaning of natural language expressions in terms of computer programs that can be evaluated against images, databases, knowledge graphs or other external resources. While resources annotated with procedural semantic representations already exist for a variety of spoken languages, such resources are still lacking entirely for signed languages. In this paper, we introduce GeoQuery-LSFB as a signed language extension to the multilingual GeoQuery corpus. Concretely, we have complemented each procedural semantic annotation from the original corpus with a corresponding French Belgian Sign Language (LSFB) expression that was phonetically transcribed from video recordings following the HamNoSys convention and annotated with French ID-glosses. The GeoQuery-LSFB corpus constitutes a substantial new resource for a low-resource language and offers for the first time the possibility to study, from an onomasialogical perspective, a signed language along a diverse variety of spoken languages.[1]

1 Introduction

Procedural semantic representations describe the meaning of natural language expressions in terms of computer programs that are compositionally structured and can be evaluated by a machine (Woods, 1967; Johnson-Laird, 1977; Woods, 1981; Winograd, 1972; Woods, 2010). A defining property of procedural semantics is that the evaluation of semantic representations involves their grounding in some kind of 'world model', which can range from a database or knowledge graph, through a quantitative or qualitative simulation, to the actual world as perceived through a robot's sensory system. The grounded and compositional nature of procedural semantic representations is of great interest to a variety of natural language understanding tasks, including database querying (Zelle and Mooney, 1996; Kwiatkowski et al., 2010; Berant et al., 2013; Liang, 2016; Dong and Lapata, 2016; Cheng et al., 2019), visual question answering and dialogue (Andreas et al., 2016; Johnson et al., 2017b; Hudson and Manning, 2019; Verheyen et al., 2023), and robot instruction (Bollini et al., 2013; Misra et al., 2016; van Trijp et al., 2024).

While a number of resources that annotate natural language expressions with procedural semantic representations already exist today, most of these resources exclusively include English data (Hemphill et al., 1990; Zelle and Mooney, 1996; Kuhlmann et al., 2004; Zettlemoyer and Collins, 2005; Chen and Mooney, 2008; Tasse and Smith, 2008; Johnson et al., 2017a; Hudson and Manning, 2019; Nevens et al., 2024). For other spoken languages, resources that come with procedural semantic annotations exist much more scarcely, but some corpora are available for German (Gross et al., 2018; Jones et al., 2012), Chinese (Lu and Ng, 2011), Spanish, Japanese and Turkish (Wong and Mooney, 2006), Greek and Thai (Jones et al., 2012), and Indonesian, Swedish and Farsi (Susanto and Lu, 2017). All non-English resources, with the exception of Gross et al. (2018), extend the original English GeoQuery corpus (Zelle and Mooney, 1996) with translations, resulting in a parallel corpus that includes a typologically rather diverse selection of languages. When it comes to signed languages however, no corpora annotated with procedural semantic representations exist to date.

In this paper, we introduce the GeoQuery-LSFB corpus as a signed language extension to the multilingual GeoQuery corpus. Concretely, this new resource complements each procedural semantic annotation from the original corpus with a corresponding French Belgian Sign Language (LSFB) expression. Based on video recordings that fea-

[1]The authors declare that this paper was written without the assistance of generative writing aids and that no AI assistants were used in any stage of the research.

ture a native LSFB signer, the expressions were annotated with French ID-glosses and phonetically transcribed following a time-aligned, multilinear extension to the Hamburg Notation System (HamNoSys) convention (Hanke, 2004). Apart from the 250 utterances in the original parallel corpus, we also present an augmented version of the corpus that covers 4519 utterances and thereby better fits today's data-intensive processing methods.

The GeoQuery-LSFB corpus constitutes a significant contribution to the linguistic resource landscape in three main respects. First of all, it is the first resource that aligns signed language expressions with procedural semantic representations. The corpus thereby facilitates research into language processing technologies that could so far only be developed for spoken languages. Second, it adds valuable corpus data to the limited resources that are currently available to support the study of the LSFB sign language. Finally, the parallel and semantically annotated nature of the corpus offers for the first time the possibility to study, from an onomasiological perspective, a signed language along a diverse variety of spoken languages.

The GeoQuery-LSFB corpus was released under the GNU General Public License 2.0 and is available for download at <see supplementary materials>[2].

2 Background and related work

2.1 French Belgian Sign Language resources

LSFB (*Langue des signes de Belgique francophone*) is the sign language used by the deaf and hard-of-hearing community within the French-speaking Community of Belgium. The number of LSFB users is estimated at 20,000 of which 4,600 are first language users[3]. As a sign language, LSFB is produced and comprehended through the visual-gestural modality. Phylogenetically, LSFB is closely related to Vlaamse Gebarentaal (VGT), the sign language of the Flemish Community of Belgium. Both languages are historically rooted in the Old French Sign Language (VLSF) and were considered under the umbrella of Belgian Sign Language until the 1970s. LSFB received recognition as an official language in the French-speaking Community of Belgium in 2003.

While significant efforts in creating linguistic resources that can be used for LSFB research, education and language technology development are currently ongoing, LSFB for now remains a low-resource language. The most extensive resource today is the LSFB Corpus (Meurant, 2015; Meurant et al., 2016), which consists of 90 hours of unscripted yet task-moderated video-recorded conversations between pairs of signers. French ID-gloss annotations (Johnston, 2008) have been made available for 25 hours of the corpus material. An overview of all ID-glosses that were used to annotate the corpus, along with French translations and isolated video recordings of the corresponding LSFB signs, has been released as the lex-LSFB lexical database (Meurant et al., 2015). The LSFB online dictionary (`https://dico.lsfb.be`) provides LSFB video translations for over 5000 French words and a tool for translating LSFB signs into French words was made available by Fink et al. (2022). The LSFB Corpus was made available to the machine learning community through the release of two datasets, one of which wraps the original corpus data (LSFB-CONT) while the other groups occurrences of the most frequently used signs (LSFB-ISOL) (Fink et al., 2021). A spoken Belgian-French counterpart to the LSFB Corpus, where participants were asked to carry out the same conversational tasks, is currently under construction (Lepeut et al., 2024).

2.2 Sign language transcription

On a high level, sign language transcription systems can be grouped into two categories. The first category transcribes each sign within a signed expression using a gloss, i.e. an identifier for the sign that essentially corresponds to the translation of the sign into a different language, typically the ambient spoken language of the community. Of particular interest are ID-glosses (Johnston, 2008, 2010), where contextually different variants of a sign, i.e. variants with the same form but a different meaning, are grouped under the same gloss. (ID-)glosses do not contain any information about the form of a sign itself and are heavily dependent on a spoken glossing language.

The second category of writing systems transcribe signed expressions phonetically or phonologically. Systems in this category do not rely on any external glossing language, but transcribe

[2]Upon acceptance of this paper, the corpus will be archived on Zenodo and will be referenced through its DOI in the paper.

[3]The overall number of LSFB users is estimated at 0.44% of the total population following Haeusler et al. (2014). The number of first language users is estimated at 0.1% of the total population following Pasikowska-Schnass (2018).

the form components of signs directly. Phonetic/phonological transcription systems for signed languages were pioneered by Stokoe (1960) for American Sign Language (ASL). The SignWriting (Sutton, 1995) and HamNoSys (Prillwitz et al., 1987; Hanke, 2004) writing systems are currently the most widely used language-agnostic transcription systems. SignWriting logographically represents manual and non-manual components of signs in 2D space using an alphabet of 652 base symbols (Sutton, 2010). HamNoSys provides a linear, phonographic representation of sign components using a more restricted inventory of just over 200 base symbols (Smith, 2013). More recently, the Typannot writing system for handshape annotation was introduced with the goal of reaching a HamNoSys-like phonographic system while achieving a SignWriting-like readability (Rébulard et al., 2018), but is still to find a more widespread adoption.

Gloss-based transcriptions and phonetic transcriptions are very different in nature, have different properties and serve different purposes. Gloss-based transcriptions are, relatively speaking, less time-consuming to annotate and their grounding in an ambient spoken language naturally supports their use in domains where the relationship between the signed and the spoken language is of significant interest, such as (machine) translation and different forms of sign language education. Phonetic transcriptions on the other hand are a valuable resource for linguistic research as they allow for a representation of signed forms that do not need to be cast into linguistic theories that were developed for spoken languages. As noted by Hodge and Crasborn (2022), it can be considered good practice in corpus development to incorporate different levels and types of annotation. In this way, the corpus becomes multipurpose and is thereby likely to better stand the test of time.

2.3 Procedural semantics

The procedural approach to semantics was introduced in the field of artificial intelligence in the 1960s and 1970s by a.o. Woods (1967), Winograd (1972) and Johnson-Laird (1977). The innovative idea behind this approach was that the meaning of a natural language expression can take the form of a program that can be evaluated by a machine. The evaluation of such a program involves its grounding in a 'world model', i.e. an external resource such as a database, knowledge graph, sensory observation, or simulation. Let us consider an example drawn from the GeoQuery resource, which will be described in more detail in the next section. The world model consists in a database of geographical facts, including the area and population of cities and states, the height of mountains, the length of rivers, and the states and cities rivers run through. Procedurally, the meaning of a question like "*How big is Texas?*" could be represented through a query that, when evaluated against the database, would return the area of the state of Texas. Importantly, the answer itself is irrelevant when it comes to representing the meaning of the question. If a new treaty is signed and the world model is consequently updated, a query that accurately captured the meaning of the question would still return the correct answer, even if the answer is no longer the same. Likewise, a program that adequately represents the meaning of the statement "*The Mississippi river runs through Iowa*" should evaluate to true in every world where Iowa is among the states the Mississippi river runs through, and evaluate to false in any other world.

The procedural approach to representing meaning subscribes to the Davidsonian view that the meaning of an utterance corresponds to its truth conditions (Davidson, 1967). A complete understanding of an utterance amounts to being able to determine its truth value in every possible world, a task that requires an exact knowledge of the conditions under which the utterance is true. In terms of procedural semantics, the programs that represent the meaning of natural language utterances thus correspond to their truth conditions. The process of evaluation then corresponds to the task of determining the truth of an utterance in a particular world[4]. In general, there are many ways in which procedural semantic programs can be formalised and in which their evaluation can be implemented, and the choice will typically depend on the task at hand and the nature of the world model that is provided (see e.g. Verheyen et al., 2023). Despite the terminology that is used, procedural semantic programs almost never take the form of a sequence of operations to execute. Instead, they state the logic underlying the computation they are supposed to represent, for example through a conjunction of predicates. As long as a procedural interpretation can be linked to their logic interpretation, the programs are consid-

[4] In Fregean terminology, the programs would correspond to the *Sinn* of an utterance, while the evaluation process would reveal its *Bedeuting* (Frege, 1892).

ered procedural semantic representations. In this paper and the associated GeoQuery-LSFB corpus, all procedural semantic representations will take the form of logic expressions that correspond to machine-evaluatable database queries.

2.4 The GeoQuery corpus

The GeoQuery corpus was originally introduced by Zelle (1995) and Zelle and Mooney (1996) as a benchmark dataset for the supervised learning of semantic parsers. In its original form, it consisted of 250 English questions[5] about US geography that were collected from undergraduate students at the Department of Computer Sciences of the University of Texas at Austin. Each question was annotated by a human expert with a logical query that would answer the question upon evaluation against the GeoBase database of US state geography (Borland International, 1988). The corpus was extended to 880 question-query pairs by Tang and Mooney (2001) and Tang (2003) using data collected from users of the GeoQuery web interface.

The logical queries annotated in the original dataset straightforwardly integrate with logic programming languages such as Prolog. In order to better accommodate semantic parsers written in other programming languages, Kate et al. (2005) introduced FunQL, a functional, variable-free query language for the GeoQuery domain, along with scripts for the automatic conversion between the annotated logical queries and their FunQL counterparts. Popescu et al. (2003) and Iyer et al. (2017) introduced a relational database schema for the Geobase database and manually annotated all GeoQuery questions with SQL queries.

The first multilingual version of the GeoQuery corpus was introduced by Wong and Mooney (2006), who provided Spanish, Japanese and Turkish translations of the original GeoQuery questions. Later, translations of all questions in the extended GeoQuery dataset were added for many other languages, in particular for Chinese by Lu and Ng (2011), for German, Greek and Thai by Jones et al. (2012) and for Indonesian, Farsi and Swedish by Susanto and Lu (2017).

Figure 1 shows a single entry in the multilingual GeoQuery corpus. The original English question "*Through which states does the Mississippi run?*" is shown on top along with its translation into the 10 other languages of the dataset. The procedural semantic representations that correspond to the questions are shown in the bottom-left part of the figure. Note that the logical query and the FunQL query are equivalent, but that the SQL query was constructed independently. A schematic representation of the 'world models' is shown in the bottom-right part of the figure. The SQL meaning representation can be evaluated against the relational database, while the GeoQuery and FunQL meaning representations can be evaluated against the Prolog factbase.

[5] We refer to the GeoQuery utterances as *questions* as they constitute requests for information to be answered. Grammatically, these requests can be expressed through interrogative sentences (e.g. '*What is the city with the smallest population?*') but also through imperative sentences (e.g. '*Give me the biggest city in Wisconsin*').

3 Data collection and transcription

The overall corpus creation task consisted in annotating the 250 utterances of the original GeoQuery corpus with corresponding LSFB expressions. Given that LSFB is not a written language, one cannot simply ask a native signer to write down LSFB translations of the English sentences. Our native LSFB informant was however experienced in signing LSFB expressions that semantically correspond to utterances written down in the ambient French language. The first stage in the corpus creation process therefore consisted in making video recordings of LSFB expressions. These recordings could then be transcribed both phonographically and in terms of ID-glosses by an expert transcriber in a second stage.

In order to optimise the workload of the LSFB informant and transcriber, the first challenge was to select the minimal number of expressions that needed to be signed, video-recorded and transcribed. We started from the 221 unique meaning representations in the corpus and reduced them to 95 schemata by replacing named entities by their higher-level entity type (<state>, <river>, <city> or <capital>). For each of these schemata, the first corresponding utterance from the GeoQuery dataset was translated into French and provided to the signer. The signer was instructed to interpret the meaning of the French utterance and sign a corresponding LSFB expression. For meaning representations that occurred multiple times in the corpus with the exact same named entities, the signer was asked to sign the same number of variants. The signed utterances were video-recorded, yielding a total of 124 recordings.

The expert transcriber annotated the recordings first with French ID-glosses and then with

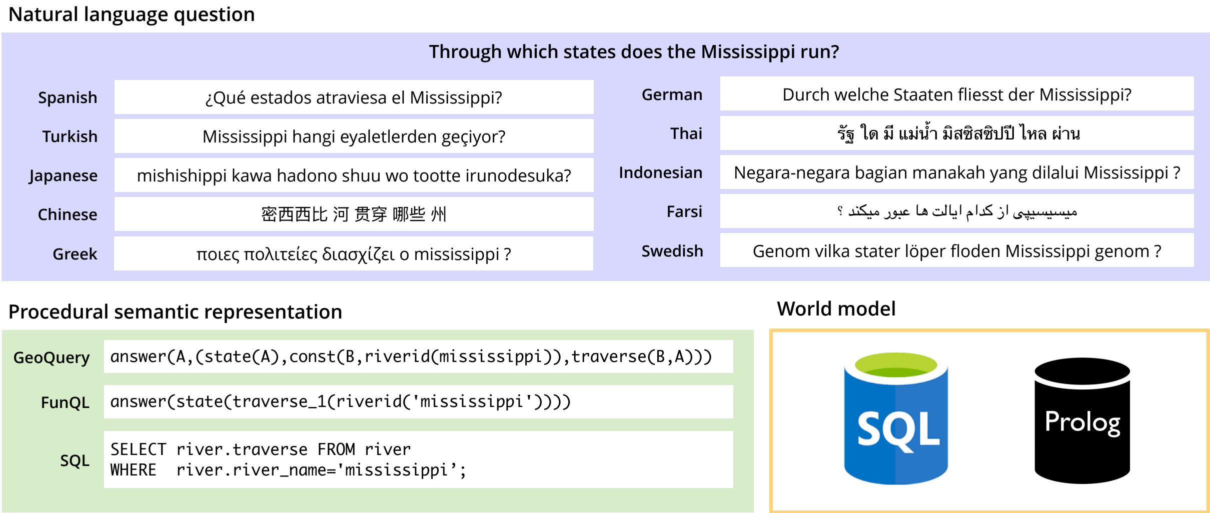

Figure 1: Schematic representation of a single entry in the multilingual GeoQuery corpus, featuring the original English question, its translation into ten other spoken languages, its procedural semantic representation in the GeoQuery, FunQL and SQL languages, and the relational and logic databases that constitute the world model.

HamNoSys-based phonographic transcriptions. ID-gloss annotation layers were created for the left and right hands using glosses from the lex-LSFB lexical database (Meurant, 2015). During the annotation process, four new namesigns were added to the database (NS:CALIFORNIE, NS:FLORIDE, NS:TEXAS and NS:MEXIQUE), as well as five new regular ID-glosses (KILOMETRE-CARRE, TAILLE, ME-DIRE, MOINE and SERRER). Depicting and pointing signs were annotated following the guidelines established by Johnston (2010, 2016). Finger-spelled named entities were transcribed as such and variants of existing ID-glosses were tagged following Johnston (2016)'s handshape categorisation.

Phonographic transcriptions were made using a time-aligned, multilinear extension to the HamNoSys convention. Concretely, a separate HamNoSys annotation layer was added for each hand. The segmentation of the layers was copied from the corresponding ID-gloss layers, temporally aligning the HamNoSys and ID-gloss transcriptions. Two-handed signs, which HamNoSys transcribes using a single linear expression, are annotated in the layer for the dominant hand, in which case the corresponding segment in the layer for the other hand is left empty.

An example transcription is shown in Figure 2 for the original utterance "*What are the high points of the states surrounding Mississippi?*". The annotation layers associated to the signer's dominant right hand are divided into eight time-aligned segments, one for each ID-gloss that was identified. The signer opens with the PALM-UP sign, a discourse marker that indicates that the speaker is thinking (Gabarró-López, 2017). This sign is followed by the DANS sign (English: IN) and the finger-spelled proper name Mississippi (FS:MISSISSIPPI, where FS stands for 'finger spelling'). Then, the IL-Y-A sign is annotated (English: THERE-IS), followed by a depicting sign DS(BENT5):ETAT+, where DS stands for 'depicting sign', BENT5 for the handshape categorised according to Johnston (2016) and + signals that the sign is repeated multiple times. Finally, the pointing sign PT:DET/LOC(1)+ is annotated, where PT stands for 'pointing sign', DET/LOC for 'determiner or location', 1 for the used handshape and + for repetition, followed by the ID-glosses HAUT (English: HIGH) and QUOI (English: WHAT). Note that the circular motion of the pointing sign is included in its HamNoSys transcription, but is reduced to 'repetition' in the ID-gloss annotation.

The annotation layers for the left hand hold three segments, two of which are ID-gloss annotations for two-handed signs (PALM-UP and DANS). For these segments, the HamNoSys transcription layer for the left hand remains empty, as the symmetry operator ¨ is used in the corresponding right-hand HamNoSys segment. The third left-hand segment concerns a depicting sign referring to a state-like entity that is again signed using Johnston (2016)'s handshape BENT5 and which coincides with three right-hand segments.

After the transcription of the 124 video record-

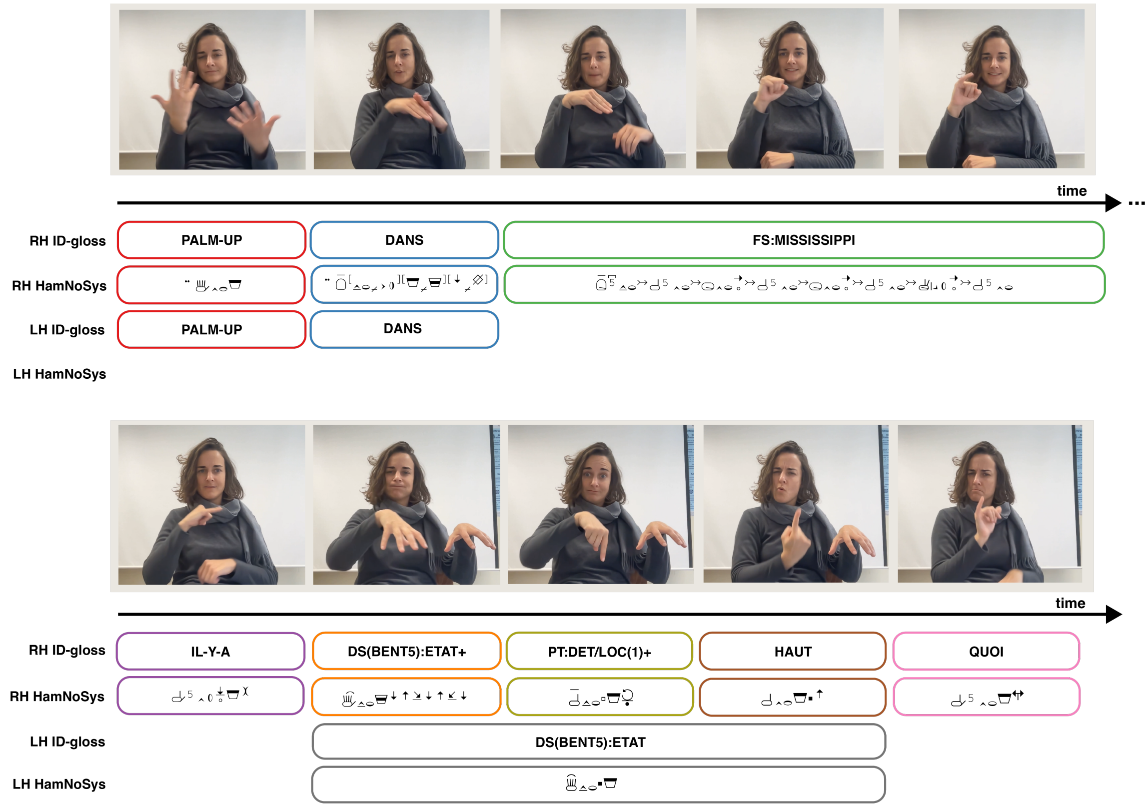

Figure 2: An example transcription of a video fragment in which the LSFB expression that corresponds to the English utterance *What are the high points of the states surrounding Mississippi?* is signed. Tiers for right-hand ID-gloss, right-hand HamNoSys transcription, left-hand ID-gloss and left-hand HamNoSys transcription are shown.

ings was completed, the resulting annotation layers were extrapolated to the 250 utterances of the original corpus. This could be done straightforwardly by substituting the named entity segments that were annotated in the recordings with segments containing the namesigns or finger-spelled transliterations of the named entities that occur at the same location in the remaining utterances. A separate, larger corpus of 4519 annotated utterances was also created by extrapolating the transcriptions of the video recordings to all named entities in the GeoBase database, enforcing the constraint that states, rivers, cities and capital cities can only be substituted with an entity of the same type.

As indicated above, the French ID-gloss annotations and HamNoSys transcriptions were contributed by a single expert transcriber. As a result, no extensive inter-annotator agreement study could be performed. While we acknowledge this as a limitation of the resource that we created, we would very much like to emphasize the difficulty of recruiting and training annotators for a low-resource sign language like LSFB. The annotator that was recruited is to the best of our knowledge the only person sufficiently proficient in both LSFB and HamNoSys to perform this task reliably. When it comes to the French ID-gloss annotations, a small-scale inter-annotator agreement study was performed in the pilot phase of the corpus creation process in order to validate its feasibility. This pilot study yielded a Cohen's kappa of 0.83 on 10 corpus entries selected to be maximally different. This result led to the green light needed to initiate the corpus creation process, but cannot be reported as such as a measure of the quality of the final resource.

All annotations were created using the ELAN software toolkit (Crasborn and Sloetjes, 2008)[6], the HamNoSys input palette (Hanke, 2021)[7] and the CWA Signing Avatars SiGML Player[8].

4 The GeoQuery-LSFB corpus release

The GeoQuery-LSFB corpus was released under the GNU General Public License 2.0 and can be

[6] Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands:https://archive.mpi.nl/tla/elan

[7] https://www.sign-lang.uni-hamburg.de/hamnosys/input/

[8] https://vhg.cmp.uea.ac.uk/tech/jas/std/

downloaded from <see supplementary materials>[9]. The release consists of the following parts:

Documentation A user guide that documents the contents of the release, especially with respect to using the corpus for research purposes. The documentation provides a detailed description of all aspects of the corpus, including the required installation procedures, the labels used during the transcription process and the data structures that were used.

Video recordings All 124 video recordings, encoded using MPEG-4 compression. The videos amount to a total length of 21 minutes and 44 seconds, which corresponds to an average of 10.35 seconds per video.

HamNoSys font A version of the Menlo font that was extended with the HamNoSys character set by Hanke (2021).

Annotation files The 124 raw ELAN annotation files in `.eaf` format, as well as their corresponding configuration files in `.psfx` format.

Corpus files The actual corpus files, which hold for each entry: a unique identifier, the original English utterance, a reference to the higher-level semantic schema of which the meaning of the utterance is an instantiation, a reference to the corresponding ELAN file and video recording (where applicable), translations of the utterance into Spanish, Turkish, Japanese (Romanji), German, Greek, Thai, Indonesian, Swedish, Chinese and Farsi, the French translation of the utterance that was provided to the LSFB signer, procedural semantic representations of the utterance in the GeoQuery, FunQL and SQL formats, and our integrated representation of the ID-gloss and HamNoSys annotations of the corresponding LSFB expression. Apart from the unique identifiers and references to the higher-level semantic schemata, ELAN files and video recordings, we contributed the French and LSFB annotations and adopted all other annotations from prior versions of the corpus as documented in Section 2.4 above. The corpus files are provided in both the JSON and XML formats, with both versions containing the exact same information.

Database files Scripts for executing and evaluating the procedural semantic representations of the utterances in GeoQuery, FunQL and SQL formats, including Iyer et al. (2017)'s script for creating and populating a relational database. Instructions on how to use the execution and evaluation scripts are included in the user guide.

[9]Upon acceptance of this paper, the corpus will be archived on Zenodo and will be referenced through its DOI in the paper.

Overall, the transcription layers of the GeoQuery-LSFB corpus for the dominant right hand contain 2681 segments, while those for the left hand hold 1258 segments. 2794 segments (70.93%) were annotated with lexicalised ID-glosses. 143 distinct lexicalised ID-glosses occur, with the most frequent ones being either generally frequent in LSFB, such as UN (English: A/ONE) and IL-Y-A (English: THERE-IS), or being related to the domain of the dataset (e.g.VILLE (English: CITY) and RIVIERE (English: RIVER)). 595 segments (15.11%) were annotated with pointing sign glosses and 380 segments (9.65%) were annotated with depicting sign glosses. The remaining 170 segments (4.32%) were annotated as finger-spelled proper names.

5 Discussion and conclusions

We have presented the GeoQuery-LSFB corpus as a linguistic resource for the low-resource French Belgian Sign Language (LSFB). Concretely, we have extended the multilingual and semantically annotated GeoQuery corpus with transcriptions of corresponding LSFB expressions. Based on video recordings featuring a native LSFB signer, an expert transcriber has annotated each expression from the original corpus with French ID-glosses (following Meurant et al., 2015; Johnston, 2016) and phonographic HamNoSys transcriptions (following Hanke, 2004). The corpus was released under the GNU General Public License 2.0 and includes (i) an extensive user guide, (ii) the video recordings that were made, (iii) the original English corpus data and procedural semantic annotations, (iv) the multilingual versions of the corpus and FunQL and SQL annotations that were previously contributed by different researchers, (v) our LSFB annotations in terms of French ID-glosses and HamNoSys transcriptions, and (vi) scripts for executing and evaluating the procedural semantic representations. A synthetically augmented version of the corpus was also included in order to meet user demand.

The GeoQuery-LSFB corpus constitutes a significant contribution to the linguistic resource landscape in three main respects. First of all, the corpus contributes valuable data to the limited pool of resources that are currently available for the linguistic study of LSFB, as well as for the development of LSFB language technologies. Such data is challenging to create, not only for financial reasons, but also given the limited number of people that both sufficiently master LSFB and have experience or interest in the creation and management of linguistic resources. Second, the GeoQuery-LSFB corpus is, to the best of our knowledge, the first corpus that aligns signed expressions with procedural semantic annotations. The corpus thereby facilitates research into language processing technologies, such as data-efficient semantic parsers or natural language interfaces, that could so far only be developed for spoken languages. Finally, the parallel and semantically annotated nature of the corpus offers for the first time the possibility to study, from an onomasiological perspective, a signed language along a diverse variety of spoken languages.

Limitations

Single annotator The LSFB transcriptions and annotations were performed by a single expert transcriber. We acknowledge this as a limitation of the resource that we created. Yet, we would like to emphasize the difficulty of recruiting and training transcribers for a low-resource sign language like LSFB. We were not able to identify a second potential transcriber sufficiently proficient in both LSFB and HamNoSys to perform the transcription task. Regarding the French ID-gloss annotations, a small-scale inter-annotator agreement study was performed in the pilot phase of the corpus creation process in order to validate its feasibility. This pilot study yielded a Cohen's kappa of 0.83 on 10 corpus entries selected to be maximally different. This result led to the green light needed to initiate the corpus creation process, but the final corpus was annotated by only one of the transcribers that was part of the pilot due to a lack of available human resources.

Limited size and domain of the corpus With its 250 and 4519 utterances, the resource that we introduce is limited in volume and restricted to a single domain. As such, it might not be suitable for many data-intensive processing methods and should not be taken as a representative sample of language use. At the same time, it is the first resource of its kind for LSFB, and for any signed language for that matter. Given the extremely limited resources available for the low-resource LSFB language, the corpus will find immediate use in LSFB-related research and to the best of our expectations in sign language research more generally.

Ethics Statement

The LSFB informant consented to the publication and distribution of the video recordings that are part of the GeoQuery-LSFB resource. Both the LSFB informant and the expert transcriber were remunerated for their work according to the official pay scales for scientific personnel employed by universities of the French community of Belgium.

References

Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. 2016. Learning to compose neural networks for question answering. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1545–1554. Association for Computational Linguistics.

Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic parsing on freebase from question-answer pairs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1533–1544.

Mario Bollini, Stefanie Tellex, Tyler Thompson, Nicholas Roy, and Daniela Rus. 2013. Interpreting and executing recipes with a cooking robot. In *Experimental Robotics: The 13th International Symposium on Experimental Robotics*, pages 481–495. Springer.

Borland International. 1988. *Turbo Prolog 2.0 Reference Guide*. Borland International, Scotts Valley, CA.

David L. Chen and Raymond J. Mooney. 2008. Learning to sportscast: a test of grounded language acquisition. In *Proceedings of the 25th International Conference on Machine learning*, pages 128–135.

Jianpeng Cheng, Siva Reddy, Vijay Saraswat, and Mirella Lapata. 2019. Learning an executable neural semantic parser. *Computational Linguistics*, 45(1):59–94.

Onno Crasborn and Han Sloetjes. 2008. Enhanced ELAN functionality for sign language corpora. In *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 39–43. European Language Resources Association (ELRA).

Donald Davidson. 1967. Truth and meaning. In *Philosophy, Language, and Artificial Intelligence: Resources for Processing Natural Language*, pages 93–111. Springer.

Li Dong and Mirella Lapata. 2016. Language to logical form with neural attention. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33–43. Association for Computational Linguistics.

Jérôme Fink, Pierre Poitier, Maxime André, Laurence Meurant, Benoît Frénay, and Anthony Cleve. 2022. Dictionnaire contextuel langue des signes belge francophone vers français.

Jérôme Fink, Benoît Frénay, Laurence Meurant, and Anthony Cleve. 2021. LSFB-CONT and LSFB-ISOL: Two new datasets for vision-based sign language recognition. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Gottlob Frege. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100(1):25–50.

Sílvia Gabarró-López. 2017. *Discourse Markers in French Belgian Sign Language (LSFB) and Catalan Sign Language (LSC): BUOYS, PALM-UP and SAME.* Ph.D. thesis, Faculty of Arts, Université de Namur, Namur, Belgium.

Stephanie Gross, Matthias Hirschmanner, Brigitte Krenn, Friedrich Neubarth, and Michael Zillich. 2018. Action Verb Corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA).

Laurcence Haeusler, Thibaud de Laval, and Charlotte Millot. 2014. Etude quantitative sur le handicap auditif à partir de l'enquête "handicap-santé". Technical Report 131, Direction de la recherche, des études, de l'évaluation et des statistiques (DREES), Paris, France.

Thomas Hanke. 2004. HamNoSys: representing sign language data in language resources and language processing contexts. In *LREC*, volume 4, pages 1–6.

Thomas Hanke. 2021. Hamnosys.

Charles T. Hemphill, John J. Godfrey, and George R. Doddington. 1990. The ATIS spoken language systems pilot corpus. In *Speech and Natural Language: Proceedings of a Workshop held at Hidden Valley*. Morgan Kaufmann.

Gabrielle Hodge and Onno Crasborn. 2022. *Good practices in annotation*, pages 46–89. Gallaudet University Press, Washington, D.C.

Drew A Hudson and Christopher D Manning. 2019. Gqa: A new dataset for real-world visual reasoning and compositional question answering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6700–6709.

Srinivasan Iyer, Ioannis Konstas, Alvin Cheung, Jayant Krishnamurthy, and Luke Zettlemoyer. 2017. Learning a neural semantic parser from user feedback. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 963–973.

Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C. Lawrence Zitnick, and Ross Girshick. 2017a. CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2901–2910, Honolulu.

Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Judy Hoffman, Li Fei-Fei, C. Lawrence Zitnick, and Ross Girshick. 2017b. Inferring and executing programs for visual reasoning. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2989–2998, Washington, D.C., USA. IEEE Computer Society.

Philip N. Johnson-Laird. 1977. Procedural semantics. *Cognition*, 5(3):189–214.

Trevor Johnston. 2008. Corpus linguistics and signed languages: no lemmata, no corpus. In *sign-lang@ LREC 2008*, pages 82–87. European Language Resources Association (ELRA).

Trevor Johnston. 2010. From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15(1):106–131.

Trevor Johnston. 2016. *Auslan Corpus Annotation Guidelines*.

Bevan Jones, Mark Johnson, and Sharon Goldwater. 2012. Semantic parsing with Bayesian tree transducers. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 488–496. Association for Computational Linguistics.

Rohit J Kate, Yuk Wah Wong, and Raymond J Mooney. 2005. Learning to transform natural to formal languages. In *Proceedings of the 20th national conference on Artificial intelligence-Volume 3*, pages 1062–1068.

Gregory Kuhlmann, Peter Stone, Raymond Mooney, and Jude Shavlik. 2004. Guiding a reinforcement learner with natural language advice: Initial results in robocup soccer. In *The AAAI-2004 workshop on supervisory control of learning and adaptive systems*. San Jose, CA.

Tom Kwiatkowski, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman. 2010. Inducing probabilistic CCG grammars from logical form with higher-order unification. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1223–1233.

Alysson Lepeut, Clara Lombart, Sébastien Vandenitte, and Laurence Meurant. 2024. Spoken and signed languages hand in hand: parallel and directly comparable corpora of french belgian sign language (lsfb) and french. *Corpora*, 19(2):241–253.

Percy Liang. 2016. Learning executable semantic parsers for natural language understanding. *Communications of the ACM*, 59(9):68–76.

Wei Lu and Hwee Tou Ng. 2011. A probabilistic forest-to-string model for language generation from typed lambda calculus expressions. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1611–1622. Association for Computational Linguistics.

Laurence Meurant. 2015. *Corpus LSFB. First digital open access corpus of movies and annotations of French Belgian Sign Language (LSFB)*. LSFB-Lab, University of Namur.

Laurence Meurant, Christophe De Clerck, Sibylle Fonze, Gauthier Raes, Susana Sanchez, and Éric Bernagou. 2015. Lex-LSFB: online lexical database from and linked to the 'Corpus LSFB'.

Laurence Meurant, Aurelie Sinte, and Eric Bernagou. 2016. The french belgian sign language corpus a user-friendly searchable online corpus. In *7th workshop on the Representation and Processing of Sign Languages: Corpus Mining (LREC 2016)*, pages 167–174.

Dipendra K Misra, Jaeyong Sung, Kevin Lee, and Ashutosh Saxena. 2016. Tell me Dave: Context-sensitive grounding of natural language to manipulation instructions. *The International Journal of Robotics Research*, 35(1-3):281–300.

Jens Nevens, Robin De Haes, Rachel Ringe, Mihai Pomarlan, Robert Porzel, Katrien Beuls, and Paul Van Eecke. 2024. A benchmark for recipe understanding in artificial agents. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pages 22–42.

Magdalena Pasikowska-Schnass. 2018. Sign languages in the eu. Technical Report PE 625.196, European Parliamentary Research Service (EPRS), Brussels, Belgium.

Ana-Maria Popescu, Oren Etzioni, and Henry Kautz. 2003. Towards a theory of natural language interfaces to databases. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 149–157.

Siegmund Prillwitz, Regina Leven, Heiko Zienert, Thomas Hanke, Jan Henning, et al. 1987. HamNoSys. Hamburg Notation System for sign language. an introduction. *Zentrum für Deutsche Gebärdensprache*.

Morgane Rébulard, Claire Danet, Adrien Contesse, Claudia S. Bianchini, Jean-François Dauphin, Léa Chèvrefils, Patrick Doan, and Dominique Boutet. 2018. TYPANNOT, a new glyphic system to annotate handshapes in any sign languages. In *8th Conference of the International Society of Gesture Studies "Gesture and diversity" (ISGS8)*, pages 139–140, Cape Town, South Africa. Heather Brooks.

Robert Smith. 2013. Hamnosys 4.0 user guide. Technical report, Institute of Technology Blanchardstown, Blanchardstown, Ireland.

William C. Stokoe. 1960. Sign language structure: An outline of the visual communication system of the american deaf. *Studies in Linguistics: Occasional Papers*, 8.

Raymond Hendy Susanto and Wei Lu. 2017. Semantic parsing with neural hybrid trees. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pages 3309–3315.

Valerie Sutton. 1995. Lessons in signwriting: textbook and workbook. *The Center for Sutton Movement Writing, Inc.*

Valerie Sutton. 2010. *The SignWriting Alphabet*.

Lappoon R. Tang. 2003. *Integrating Top-down and Bottom-up Approaches in Inductive Logic Programming: Applications in Natural Language Processing and Relational Data Mining*. Ph.D. thesis, Department of Computer Sciences, University of Texas, Austin, TX.

Lappoon R Tang and Raymond J Mooney. 2001. Using multiple clause constructors in inductive logic programming for semantic parsing. In *European Conference on Machine Learning*, pages 466–477. Springer.

Dan Tasse and Noah A. Smith. 2008. SOUR CREAM: Toward semantic processing of recipes. Technical Report CMU-LTI-08-005, Carnegie Mellon University, Pittsburgh, PA, USA.

Remi van Trijp, Katrien Beuls, and Paul Van Eecke. 2024. The proof is in the almond cookies. In Luc Steels and Robert Porzel, editors, *Narrative-based Understanding of Everyday Activities: A Cookbook*, pages 59–77. Venice International University, Venice, Italy.

Lara Verheyen, Jérôme Botoko Ekila, Jens Nevens, Paul Van Eecke, and Katrien Beuls. 2023. Neuro-symbolic procedural semantics for reasoning-intensive visual dialogue tasks. In *Proceedings of the 26th European Conference on Artificial Intelligence (ECAI 2023)*, pages 2419–2426, Kraków.

Terry Winograd. 1972. Understanding natural language. *Cognitive Psychology*, 3(1):1–191.

Yuk Wah Wong and Raymond J. Mooney. 2006. Learning for semantic parsing with statistical machine translation. In *Proceedings of the 2006 main conference on Human Language Technology Conference of the North American Chapter of the Association of*

Computational Linguistics. Association for Computational Linguistics.

William A. Woods. 1967. *Semantics for a question-answering system*. Ph.D. thesis, Harvard University, Cambridge, MA.

William A. Woods. 1981. Procedural semantics as a theory of meaning. In Aravind K. Joshi, Bruce L. Webber, and Ivan A. Sag, editors, *Elements of Discourse Understanding*, pages 300–334. Cambridge University Press.

William A. Woods. 2010. ACL lifetime achievement award: The right tools: Reflections on computation and language. *Computational Linguistics*, 36(4):601–630.

John M. Zelle and Raymond J. Mooney. 1996. Learning to parse database queries using inductive logic programming. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence - Volume 2*, pages 1050–1055, Washington, D.C., USA. AAAI Press.

John Marvin Zelle. 1995. *Using inductive logic programming to automate the construction of natural language parsers*. Ph.D. thesis, The University of Texas at Austin.

Luke Zettlemoyer and Michael Collins. 2005. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, pages 658–666. AUAI Press.