

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

The French Belgian Sign Language Corpus. A User-Friendly Corpus Searchable Online

Meurant, Laurence; Sinte, Aurélie

Published in:

Proceedings of the 7th workshop on the Representation and Processing of Sign Languages:Corpus Mining

Publication date:

2016

Document Version

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for pulished version (HARVARD):

Meurant, L & Sinte, A 2016, The French Belgian Sign Language Corpus. A User-Friendly Corpus Searchable Online. in *Proceedings of the 7th workshop on the Representation and Processing of Sign Languages:Corpus Mining: LREC 2016*. pp. 166-174, LREC 2016 (10th Language Resources and Evaluation Conference), Portoroz, Slovenia, 23/05/16. <http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-SignLanguage_Proceedings.pdf>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The French Belgian Sign Language Corpus A User-Friendly Searchable Online Corpus

Laurence Meurant, Aurélie Sinte, Eric Bernagou

FRS-FNRS, University of Namur

Rue de Bruxelles, 61 - 5000 Namur - Belgium

laurence.meurant@unamur.be, aurelie.sinte@unamur.be, eric.bernagou@unamur.be

Abstract

This paper presents the first large-scale corpus of French Belgian Sign Language (LSFB) available via an open access website (www.corpus-lsfb.be). Visitors can search within the data and the metadata. Various tools allow the users to find sign language video clips by searching through the annotations and the lexical database, and to filter the data by signer, by region, by task or by keyword. The website includes a lexicon linked to an online LSFb dictionary.

Keywords: French Belgian Sign Language, searchable corpus, lexical database.

1. The LSFb corpus

1.1. The project

In Brussels and Wallonia, i.e. the French-speaking part of Belgium, significant advances have recently been made of the development of LSFb. It was officially recognised in 2003 by the Parliament of the Communauté française de Belgique. Since 2000, a bilingual (LSFB-French) education programme has been developed in Namur that includes deaf pupils within ordinary classes (Ghesquière et al., 2015; Ghesquière et Meurant, 2016). The first MA in Translation and Interpreting in LSFb-French opened in September 2014. And since the early 2000s, linguistic research has been conducted at the University of Namur. But as is the case for most sign languages, French Belgian Sign Language (LSFB) remains a less-resourced language. Until very recently, only small sets of recordings existed, most of which were private archives. A large-scale searchable corpus of videotaped data, documented by metadata about the signers and the tasks produced was direly needed.

In December 2015, the LSFb corpus website was launched, containing the results of the *Corpus LSFb* project conducted at the University of Namur between 2012 and 2015. The first aim of the project was to collect data for linguistic research on LSFb. However, we kept a close watch to ensure that the LSFb Corpus was also a useful tool for teachers, interpreters and students, and that it safeguards the cultural and linguistic heritage of the (French Belgian) Deaf Community. The corpus is available as an open access website¹ containing video data, annotations, translations and metadata. Several search options allow users to browse the data.

1.2. The setting and the technical equipment

The video recordings were collected in the fully equipped studio of the LSFb-Laboratory². The participants were invited to come in pairs, and were guided by a deaf moderator. All were seated on chairs without armrests. Three JVC

Pro HD 3 CCD cameras recorded the participants: one for an upper body view of each informant (Cam 1 and 2 in Figure 1), and one for a wide-shot of both of them (Cam 3 in Figure 1). Additionally, a Sony DV Handycam was used to record the moderator (Cam 4 in Figure 1). The positions of the participants and the cameras are illustrated in Figure 1.

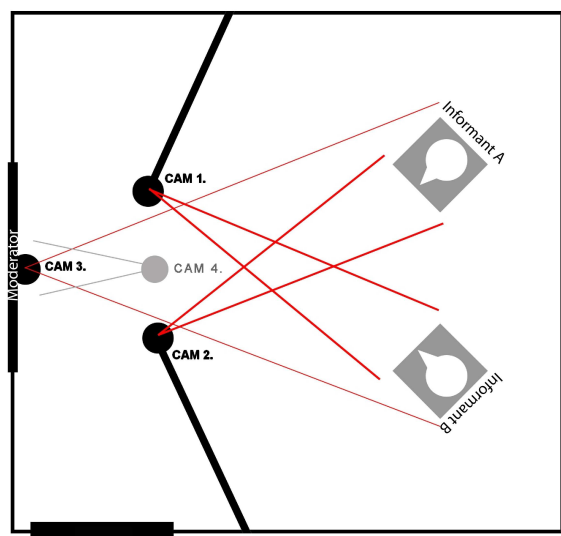


Figure 1: Positions of the participants and the cameras

1.3. The recording and the editing of the data

The data were recorded in full HD resolution (1920x1980 pixels), at 50 frames per second. Then they were edited and compressed with the software EDIUS. The edited files were exported in two sizes, both in .mp4 at 50 frames per second: in 1920x1080 pixels and in 720x576 pixels. These compression formats appeared to be the most convenient for using the videos in ELAN³ either on PC or on Mac. The video files are cut and named according to the recording sessions (1 to 50), the tasks (1 to 19) and the camera shots (B: upper body shots, L: wide shot, M: moderator shot). In

¹<https://www.corpus-lsfb.be>

²<https://www.unamur.be/lettres/romanes/lisfb-lab>

³<http://tla.mpi.nl/tools/tla-tools/elan/>

other words, one video file refers to a whole task performed by the two participants of a particular session.

During each session and each task, the cameras recorded all the exchanges: the explanations and instructions given by the moderator before the task and at times during the task, for example if the participants asked a question or needed more explanations, the times when the participants look at a piece of material, memorise it and prepare their production, as well as the exchanges between the two participants performing the task. However, at the editing stage we separated the exchanges between the participants that we considered as the ‘answer’ part of the data from the rest of the productions, called the ‘questions’. For each task, the video files are organised in one question file (Q) and three synchronised answer files (A).

- The Q file shows the four shots at the same time, i.e. the two upper body views, the wide shot on both participants and the moderator view, corresponding to the ‘questions’ sequences. The Q file is comprised of the succession of initial instructions and explanations, followed by, in chronological order, all the sequences not considered part of the data, namely the moderator’s interventions and the participants preparations. The succession of these sequences is signalled by a change in the colour of the cross shape that separates the four shots on the screen. The colours are always used in the same order: black, red, yellow, green, orange, turquoise, blue and white.
- The A files contain the actual exchanges: the participants sign to each other while looking at each other. One A file corresponds to one camera shot: L is the wide shot, M is the moderator shot, S00A-B is the upper body shot showing participant 00A and S00B-B is the upper body shot showing participant 00B. When the exchanges between the participants are interrupted, these interruptions are signalled by a fully-coloured screen. The colours used and their order correspond to what has been established for the Q files. In this way, it is possible to link the different sequences of the Q file with their context in the A files of the same task.

1.4. The signers

The objective of the Corpus LSFb project was to collect a representative sample of the LSFb signs currently in use in Brussels and Wallonia. Yet due to the demographic features of deafness, only 5% of the signers are native signers, and amongst them only a small number have parents who are native signers themselves (Van Herreweghe and Vermeerbergen, 2012). A great number of the signers have acquired LSFb during the first years of their schooling in deaf schools; we considered them as near-native signers. For others, LSFb became their everyday language after age 7 or even during adolescence; we considered them as late signers. The LSFb corpus includes all these three profiles: 30% are native signers, 26% are near-native and 44% are



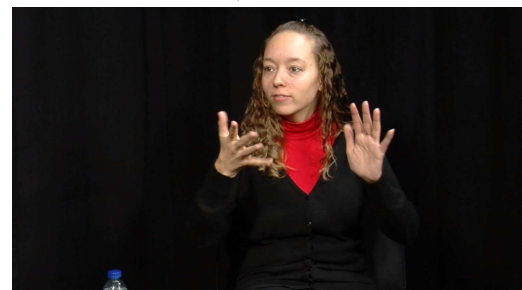
Q file (with black cross between the 4 views)



A file, L shot



A file, S00A-B shot



A file, S00B-B shot

Figure 2: This figure shows the appearance of the question file and 3 answer files

late signers. Other variables related to the signers are represented in the data: the regional variants, the variants related to the gender and to the age of the signers. 57% of the signers are women and 43% are men. They represent four age groups: 18-25 (17%), 26-45 (49%), 46-65 (18%) and 66 + (16%), and they range from 18 to 95 years old.

100 signers participated in the data collection out of a total of an estimated number of 4,000 signers in the French-speaking part of Belgium. When it came to setting up the pairs of signers for each recording session, we used the following criteria in descending hierarchical order: the similarity in terms of linguistic profile, in terms of regional variant, in terms of age and in terms of gender.

Before starting the recordings, the moderator presented the issues related to their participation and the recording of their image to the participants. Then, they were invited to sign an informed consent form and to give their agreement for the use of their data for three purposes and the related type of distribution: for research, for education and training, and for the conservation of linguistic and cultural heritage. At this point, the participants were informed that they would be allowed to confirm or restrict their agreement afterwards, namely after having viewed their productions. Indeed, after the recording, each participant received a DVD with the video files in which they appeared. Each one was asked to confirm their agreement or to specify (with time codes) the clips they wanted to censure. Only a small proportion of the participants asked for some changes, and most accepted the open access distribution of their productions for the benefit of conserving LSFB heritage.

1.5. The tasks

The 100 signers were invited in pairs which means that, from 2013 to 2015, 50 sessions were recorded. Each session lasted between 4h and 6h, for a mean length of 3 hours of edited video per session. The signers were asked by a deaf moderator to perform 19 tasks: telling stories, memories and jokes; explaining maps, routes and pictures; explaining their name sign, their hobby; comparing, arguing, classifying objects and symbols; talking about sign language and the deaf community. The questions the participants were asked were in part inspired by corpus projects from other sign languages such as Australian Sign Language (Auslan), Sign Language of the Netherlands (NGT), German Sign Language (DGS) and Flemish Sign Language (VGT). Several tasks are quite similar to those projects in order to make comparative studies possible. For example, tasks number 2 and 3 (see Figure 1.5.) are widespread among the other Sign Language corpora, but in particular in the VGT and the DGS corpora. Task number 8 contains productions based on the same material (a village map) as tasks from the VGT and the DGS corpora. Tasks number 11 and 12 include productions from the famous ‘Horse’ and ‘Frog’ stories.

The tasks cover various genres such as narratives, explanations, descriptions, argumentations and discussions. The signers are invited to talk about the deaf community, in order to document some of its specific features: the first encounter of a deaf adult, important family celebrations, school life, relationships with hearing people etc., and about a variety of non-deaf issues and topics, in order to provide a wide range of lexicon. Figure 1.5. describes the topic of each task and figure 1.5. provides an overview of some pictures used in order to support the dialogues.

1.6. Annotations and translations

One ELAN annotation file (.eaf) has been created for each task and linked to the four synchronised video answer files, namely to the four available views on the participants (the three illustrated in figure 2, and the moderator shot). The annotation process was carried out by deaf annotators

- 1 Information to complete metadata files (age, school, family, etc.)
(Because of confidentiality, this task is not available on the website.)
- 2 Explaining the name sign of both signers
- 3 Telling a childhood memory
- 4 Explaining the benefits and disadvantages of being deaf or hearing
- 5 Explaining what ‘signing well’ means
- 6 Talking about the influence of emotions on sign language
- 7 Describing a procedure such as assembly instruction of a piece of furniture or a recipe
- 8 Describing a map or a route from any given starting point to a destination
- 9 Explaining a picture (what is special about it or what issues does it raise)
- 10 Arguing about polemic or shocking topics
(general subjects such as verbal abuse, anorexia or gay marriage)
- 11 Telling a short story: joke, comic strip or short cartoon
- 12 Telling a long story: Where are you frog? or Paperman (cartoon)
- 13 Playing a role-playing game: ‘Imagine you meet a minister and you have to convince him of... (deaf community topics)’
- 14 Talking about LSFB variations: ‘Do you easily understand young/old signers, interpreters, signers from other regions?’ ‘What are the differences?’
- 15 Talking about a hobby or a job, the material used, the way to proceed, rules, etc.
- 16 Description of drawn faces
- 17 Classifying pictures and explaining criteria
- 18 Explaining differences and similarities of objects and tools
- 19 Conclusion: discussing the activities of the day, the tasks, etc. with the moderator

Figure 3: Descriptions of the tasks

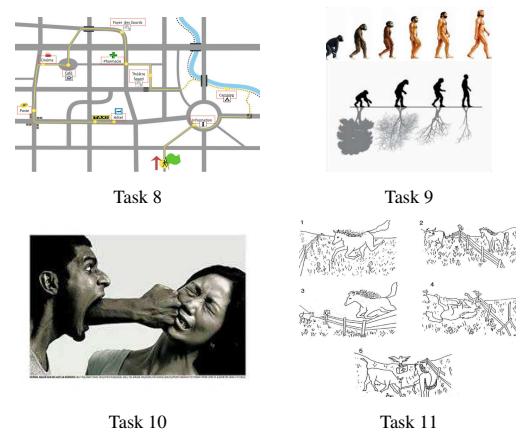


Figure 4: Sample of material

in ELAN (Sinte et al., 2015), on the basis of the Auslan Corpus Annotation Guidelines (Johnston 2015). More than 12 hours of videos are now annotated, sign by sign and hand by hand⁴, with ID-glosses (Johnston 2010). This means that all the phonological and morphological variants of a lexeme have been subsumed under the same gloss or lemma. At the time of writing this paper, 98,200 tokens out of the 104,000 annotated tokens are fully lexicalised signs. In this first step of the annotation process, the partially lexicalised signs have only been identified as such by the label ‘DS’ which stands for ‘depicting sign’ (Johnston 2015). In the ongoing second phase of annotation, a short semantic description of the partially lexicalised signs is

⁴Separate annotation tiers have been created for the right hand and for the left hand of each signer.

added within each annotation. The annotation files (.eaf) are linked and time aligned to the corpus website so that the annotations are directly visible by any user. Figure 5 shows the website interface of the video viewers and the annotations.

Among the 12 hours of annotated data, 2.5 hours have been translated so far (2,400 sentences) by a multidisciplinary team. Interpreters, linguists, bilingual teachers and deaf people took part in a workshop we organised at the University of Namur, from March to July 2015. In collaboration with Alain Bacci (Interpretis, Toulouse), who was invited as an expert for the whole workshop⁵, we established the main keystones for the translation project. The translations are target oriented: the text is produced in the most natural French possible, ensuring the French lexicon and syntactic structure reflects the influence of LSFb as little as possible. The oral features that characterise the LSFb semi-spontaneous conversations have been translated into French text in the same way as magazines do when transcribing an interview. The translation process includes three main stages. First, the translators work alone, regularly requesting the help of a deaf LSFb expert in order to clarify any point of uncertainty. Second, a referee checks the French translation and suggests corrections or modifications to the translator if needed. And third, after the modifications have been inserted, the text is validated. To date, all the available translations have been validated by Alain Bacci. Within the Corpus LSFb website, the translations may be shown at the same time as the video is playing.

Apart from this basic annotation of the data, i.e. the ID-glosses and the French translations, a great number of annotations are currently being added, mostly in connection with ongoing PhD theses that are using the data. These annotations include:

- a segmentation into Basic Discourse Units (Degand and Simon, 2005; Gabarró-López and Meurant, this volume);
- the identification and qualification of buoys and discourse markers (Gabarró-López, 2015; Gabarró-López and Meurant, 2016);
- the identification of fluency and disfluency markers (Gotz, 2013) as pauses and holds, palm-up signs, repetitions, eye gaze directions, etc. (Notarrigo and Meurant, 2016);
- the identification of phonetic phenomena such as weak hand lowering in symmetrical signs and lowering of forehead signs (Paligot et al., 2016; Paligot and Meurant, 2016);
- the identification and qualification of rephrasing structures and their relation to the original phrasing (Meurant and Sinte, 2016).

⁵Because of the lack of trained LSFb-French (the first MA started in September 2014, see section 1.1), we choose to work in collaboration with a trained of French Sign Language-French translator, who is also a trainer.

These annotations will be made available for researchers on the website within the ELAN annotation files of the concerned videos.

2. The lexical database

At the beginning of the annotation process in ELAN, the annotators needed an online lexical database to gather the ID-glosses encountered in the data, to associate each ID-gloss with a video of the sign, and to share them with the other annotators of the team. A specific web tool was developed in order to solve this issue: the Lex-LSFB.

2.1. Lex-LSFB

The tool was developed in the PHP language and the MySQL system, both being very popular and designed for web applications. Thanks to the Lex-LSFB tool, the annotators could (and still do) add or edit glosses in the database and add meta-information about each entry, such as an animated gif file which shows the sign in a video-like form for correct identification, and the possible phonological variants of the sign.

In order to link the web application to ELAN, we only had to export the lexical database as an XML file with the same features as the ‘external controlled vocabulary’ files processed by ELAN (the structure was copied from a controlled vocabulary created in ELAN). The exported file is referred to by an URL which is copied in the ad hoc field of the ‘external controlled vocabulary’ in ELAN. The controlled vocabulary is updated at every start of ELAN. If the .ecv file is updated, users must restart ELAN so that the program takes the last update into account. This tool makes the annotation work easier for the annotation team. We can also guess that it significantly improved the quality and the reliability of the annotation work.

This lexical tool is now available on the corpus website to any visitor without any hard installation and without any system requirement. It takes the form of a searchable directory containing all the glosses used in the annotation files, accompanied by keywords, i.e. possible French translations of the sign and by a video-like view of the sign in isolation.

2.2. Link to the LSFb dictionary

The web interface also includes, when possible, a link between each lexical entry and the corresponding sign within the external online dictionary of LSFb developed by the LSFb Association⁶ in collaboration with the LSFb-Laboratory of the University of Namur. For each sign, the dictionary provides a definition and examples in LSFb, etymological information when available, as well as information about the regional distribution of the sign, links towards regional variants, homonyms and synonyms. The collaboration between the Corpus LSFb project and the dictionary aims to gradually supply the dictionary, currently based on limited monological data, with the information extracted from the dialogical and wider data of the

⁶<http://dicto.lsfb.be/dico>

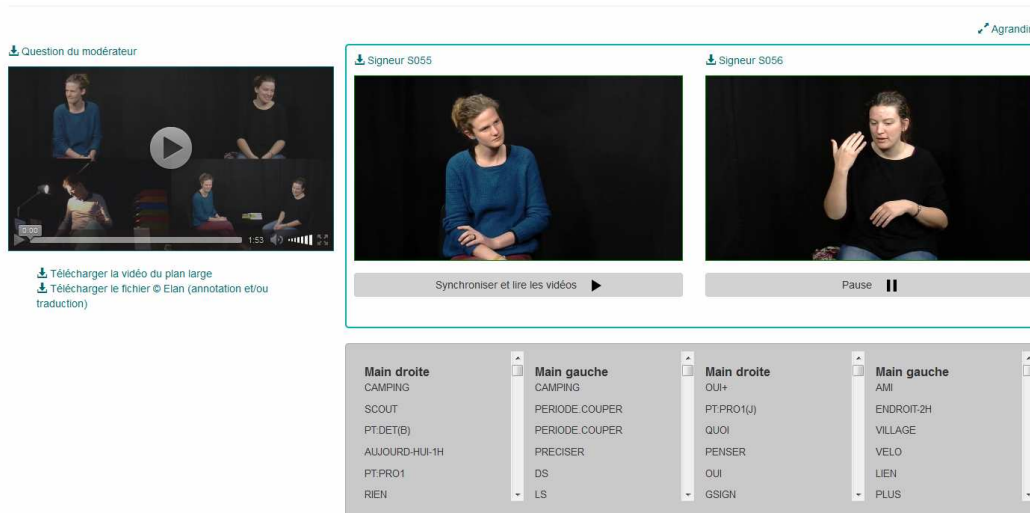


Figure 5: On the left: question file (with the 4 shots combined). On the right: single shot for each signer, as it appears on the website

LSFB corpus, presumably more representative of the language in use. Figure 6 and 7 illustrate the link between the lexical database of the corpus and the online external dictionary.

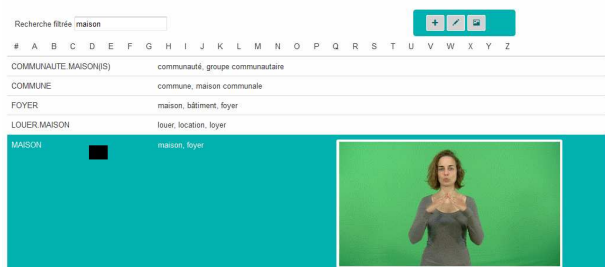


Figure 6: Lexical database and icon (a small black book) symbolising the link to the external dictionary

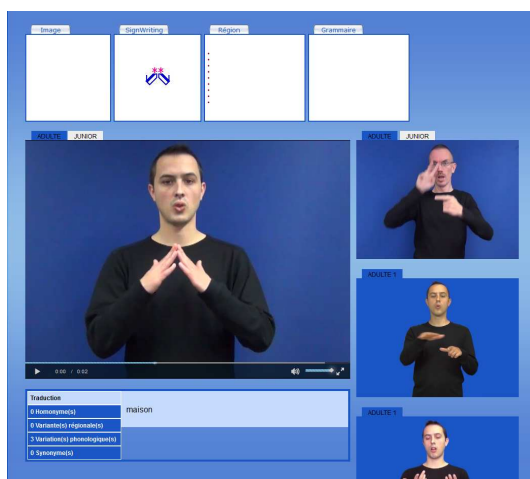


Figure 7: Online external LSFBSign dictionary

3. The website

The corpus LSFBSign website (www.lsfbsign-corpus.be) provides access to the whole content of the corpus. This includes the data (4 videos for each of the 19 tasks and for the 50 recording sessions), the metadata about signers (gender, age, profile and LSFBSign variant used) and the tasks (description and elicitation materials), the annotations and the translations. The videos showing each signer from each pair are synchronised online. All videos and .eaf ELAN files can be downloaded. The annotations and the translations can be displayed in real time while the videos are playing.

3.1. Licensing

The conditions of use of the LSFBSign Corpus website follow the BY-NC-SA Creative Commons conditions (<http://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>). Which means:

1. Attribution (BY) - Any use of the corpus has to refer to the website www.corpus-lsfbsign.be and have to cite “Meurant, L. 2015. Corpus LSFBSign. First digital open access corpus of movies and annotations of French Belgian Sign Language (LSFBSign). LSFBSign-Lab, University of Namur. URL: <http://www.corpus-lsfbsign.be>”
2. Non commercial (NC) - The website is only for non-commercial uses.
3. Share alike (SA) - Any use has to be performed under the same sharing conditions (using the same Creative Commons licence).

By these conditions, we aim to encourage the use of the data by researchers, professionals and the general public.

3.2. Search options

Users can choose the way they want to browse and query the data and metadata. The ‘Corpus’ tab offers three kinds of entries (search, free consultation, consultation by signer)

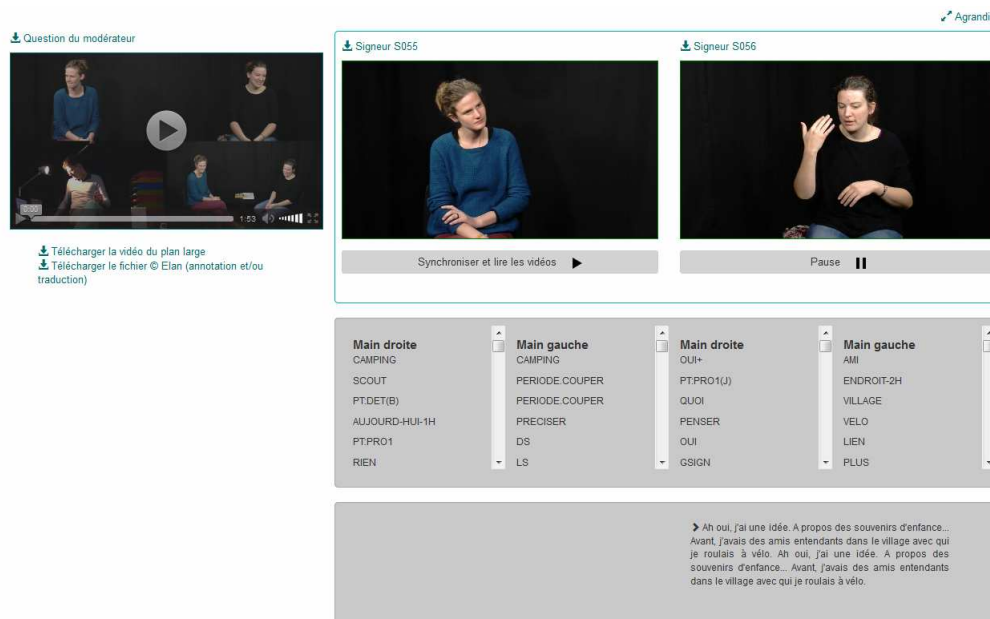


Figure 8: On the left: question file (with the 4 shots combined). On the right: single shot for each signer, as it appears on the website with annotations and translation.

and a demo that explains the general content of the website. The ‘search’ option leads the users to a page with six different ways to find videos (see figure 9):



Figure 9: The search page of the website, showing the various ways to search through the data.

1. Signer’s profile – The search by signers profile includes 4 different filters that can be combined together: gender, age, linguistic profile (native, near-native or late learners) and regional variant.
2. Regional variants – A map of Belgium allows the users to choose one region. The results give all the sessions in which signers from this region appear.

3. Descriptors – Five filters enable a query to be made regarding the content of the videos, which means the subjects covered by the discussions as well as some linguistic features. Three descriptors are related to deaf culture: school, humour and name-sign. Three descriptors concern grammatical features of the language: classifiers, use of space and iconicity. Then, three descriptors are available to search videos containing discussions about the relationship of the signers to their language: the impact of emotions on their language, the topic of the norm, and the topic of language variations. Seven descriptors relate to the topic of the discussion: childhood, family, stories, hobbies, societal issues, deaf issues and memories. Finally, eight different discourse genres are available as descriptors as well: argumentation, conversation, description, discussion, explanation, fictive narrative, life story and procedure.

4. ID-gloss or keyword – It is also possible to find video clips by searching through the ID-glosses and the content of the lexical database. When querying for a gloss, the users obtain all the videos (and the time codes) in which the gloss appears.

5. Visual material – Pretty soon, it will be possible to find videos by means of the material used for the tasks.

6. The last choice enables only the list of videos that have been annotated and/or translated so far to be displayed.

3.3. Personalised session

Users can choose simply to visit the website or to create an account. The profile of the account varies in line with the user’s profession: public, professional (teacher, interpreter) or researcher. Without an account, a visitor

to the site cannot access the videos. The public profile allows users to see the data (apart from the videos censored by the participants concerned) but not the metadata. The professional profile enables the user to view data (apart from the videos censored by the participants concerned) and some metadata; and finally, the researcher profile provides access to all data and metadata.

Each registered user accesses the corpus on their own session. This allows them to add personal comments (linked to a specific video), to tag videos as favorites and to choose some parameters such as the automatic display of annotations and/or translation, the main colour of the website, etc. At the time of writing this paper, the website is available in French and in LSF (see Figure 10). All the videos are made available for people with Usher syndrome. In July, the interface will also be available in English and International Signs. Three months after the launch of the website, 237 accounts have been created: 163 with a public profile, 56 with a professional profile (LSF teachers, interpreters and interpreting trainers) and 5 with a researcher profile. The 13 remaining accounts belongs to the LSF-Laboratory team members and the administrators of the web site.



Figure 10: Bilingual (LSF and French) interface of the website.

4. Conclusion

Collecting and sharing the LSF corpus led us to develop new tools that are expected to facilitate the study of LSF and to foster the use of corpus data by teachers, interpreters and students. Three of them are worth mentioning, because of their innovative status within the field.

1. The online lexical database linked to ELAN facilitates the annotation process.
2. The annotations and the translations, time aligned with the videos, are a real asset for user-friendly web browsing of the LSF corpus.
3. The numerous search options for browsing the data (videos), the metadata (signers and tasks) and the annotations make the LSF a very useful tool for corpus mining.

Of course, the availability of the data opens new perspectives in the development of corpus linguistic research on LSF. The major illustration of this change is the three PhD theses in preparation on LSF discourse: on

fluency and disfluency markers, on the impact of genre variation on phonetic variation and on buoys and discourse markers. But such data also open the possibility to conduct cross-linguistic studies, and in particular to investigate whether VGT and LSF, considered as one ‘Belgian Sign Language’ until very recently, have evolved as two variants of the same original language, or as two different languages.

In the short term, we consider the LSF corpus as the first step for building parallel corpora to be used with the methodology of corpus-based contrastive linguistics between LSF and French. For this purpose, we are currently collecting spoken French data on the basis of the same tasks as the ones used for the LSF corpus and in the same conditions (e.g. in a studio, with pairs of participants, a moderator, and 4 cameras). These next steps aim at answering the various issues faced by all deaf learners of French, and in particular the ones registered in the bilingual educational program in Namur, as well as their teachers, on the one hand, and by the interpreters and the interpreting trainers, on the other hand, in terms of comparison between LSF and French.

5. Acknowledgements

The LSF corpus and the website are the result of the *Corpus LSF* project funded by a three-year Incentive Grant for Scientific Research (M.I.S. F.4505.12) awarded by the F.R.S.-FNRS to Laurence Meurant. We acknowledge all the signers who participated to this project.

6. Bibliographical References

- Crasborn, O. and Zwitserlood, I. (2008). The corpus NGT: an online corpus for professionals and laymen. In O. Crasborn, et al., editors, *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*, pages 44–49. ELDA.
- Crasborn, O. and Zwitserlood, I. and Ros, J. (2008). *The Corpus NGT. An open access digital corpus of movies with annotations of Sign Language of the Netherlands*. Centre for Language Studies, Radboud University Nijmegen, ISLRN 175-346-174-413-3.
- Degand, L. and Simon, A. C. (2005). Minimal Discourse Units: Can we define them, and why should we? In *Proceedings of SEM-05. Connectors, discourse framing and discourse structure: from corpus-based and experimental analyses to discourse theories*, pages 65–74, Biarritz.
- Gabarró-López, S., Meurant, L., and Barberá, Altamira, G. (2016). *Digging into buoys: their use across genres and their status in signed discourse*. Poster presented at TISLR 12 - 12th International Conference on Theoretical Issues in Sign Language Research, Melbourne, Australia.
- Gabarró-López, S. (2016). *When "same" means something different: the sign SAME as a discourse relational device in two sign languages*. Paper presented at 3rd International Conference on Linguistic and Psycholinguistic

- tic Approaches to Text Structuring (LPTS 2016), Valencia, Espagne.
- Ghesquière, M. and Meurant, L. (2016). L'envers de la broderie. une pédagogie bilingue langue des signes de Belgique francophone - français. *Glottopol*, 27:95–111.
- Ghesquière, M., de Halleux, C., and Meurant, L. (2015). Bilingual education by immersion in Namur, Belgium: Principles and pedagogic issues. In *Proceedings of the 22nd International Congress on the Education of the Deaf*.
- Gotz, S. (2013). Fluency in native and nonnative english speech. *Studies in Corpus Linguistics*, 53.
- Trevor Johnston. (2004-2007). *Corpus of grammar and discourse strategies of deaf native users of Auslan (Australian Sign Language)*. Macquarie University.
- 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.
- Johnston, T. (2015). *Auslan Corpus Annotation Guidelines. Guidelines*, Macquarie University; La Trobe University, Sydney; Melbourne.
- Meurant, L. and Sinte, A. (2013). Towards a corpus of French Belgian Sign Language (LSFB) discourses. In C. Bolly et al., editors, *Text-Structuring: Across the Line of Speech and Writing Variation*, pages 199–212.
- Meurant, L. and Sinte, A. (2016). La reformulation en langue des signes de Belgique francophone (LSFB). analyse dans un corpus de trois types de discours: narration, explication et conversation. *L'Information grammaticale*, 149:33–45.
- Meurant, L., Cleve, A., and Crasborn, O. (2016). Using sign language corpora as bilingual corpora for data mining: Contrastive linguistics and computer-assisted annotation. volume *Proceedings of the 7th workshop on the Representation and Processing of Sign Languages: Corpus Mining : LREC 2016*.
- 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.
- Notarrigo, I. and Meurant, L. (2015). *Markers of (dis)fluency across signers' profiles in French Belgian Sign Language (LSFB). A comparative analysis between Native, Near-Native and Late Signers*. Poster presented at 2nd International Conference on Sign Language Acquisition, Amsterdam, Pays-Bas.
- Paligot, A. and Meurant, L. (2016). *Weak Hand Lowering across signing styles of French Belgian Sign Language (LSFB)*. Poster presented at TISLR 12 - 12th International Conference on Theoretical Issues in Sign Language Research, Melbourne, Australia.
- Paligot, A., Van.Der.Kooij, E., Crasborn, O., and Bank, R. (2016). *Weak Drop in Context*. Paper presented at TISLR 12 - 12th International Conference on Theoretical Issues in Sign Language Research, Melbourne, Australia.
- Prillwitz, S. (2008). DGS Corpus Project: Development of a Corpus-Based Electronic dictionary of German Sign Language/German. In *LREC 2008 Workshop Proceedings*, volume 23.
- Christian Rathmann. (2015). *DGS Corpus Online*.
- Sinte, A., De.Clerck, C., Fonzé, S., Sanchez, S., and Meurant, L. (2015). *Corpus LSFB (French Belgian Sign Language). Current annotation conventions compared with the "Digging into Signs"'s suggestions*. Poster presented at Digging into Signs Workshop: Developing Annotation Standards for Sign Language Corpora.
- Van.Herreweghe, M. and Vermeerbergen, M. (2012). Data collection. In R. Pfau M. Steinbach et al., editors, *Sign Language. An International Handbook. Handbooks of Linguistics and Communication Science (HSK) series*, volume 37, pages 1023–1045. Mouton de Gruyter.