

# Moving Towards a Functional Approach in the Flemish Sign Language Dictionary Making Process

Caro Brosens, Sam Verstraete, Thijs Vandamme, Hannes De Durpel, Margot Janssens

Vlaams GebarentaalCentrum (VGTC)  
Sint-Jacobsmarkt 49, 2000 Antwerp, Belgium  
caro.brosens@vgtc.be, sam.verstraete@vgtc.be, thijs.vandamme@vgtc.be,  
hannes.de.durpel@vgtc.be, margot.janssens@vgtc.be

## Abstract

This article outlines the dictionary making process of the new online Flemish Sign Language dictionary launched in 2019. First some necessary background information is provided, consisting of a brief history of Flemish Sign Language (VGT) lexicography. Then the three phases in the development of the renewed dictionary of VGT are explored: (i) user research, (ii) data-cleaning and modeling, and (iii) innovations. More than wanting to project a report of lexicographic research on a website, the goal was to make the new dictionary a practical, user-friendly reference tool that meets the needs, expectations, and skills of the dictionary users. Since 2017, VGTC has been using Signbank, an electronic database specifically developed to compile and manage lexicographic data for sign languages. Bringing together all this raw data inadvertently led to inconsistencies and small mistakes, therefore the data had to be manually revised and complemented. The VGT dictionary was mainly formally modernized, but there are also several substantive differences regarding the previous dictionary. Lastly, possible future innovations are briefly discussed. Future goals include adding definitions and sample sentences (preferably extracted from the corpus), as well as information on the etymology and common use of signs.

**Keywords:** Lexicography, Flemish Sign Language, Signbank, Sign Language Dictionary, Bilingual Dictionary, Bimodal Dictionary, Community Sourcing

## 1. Previous Lexicographic Work on VGT

When research into Flemish Sign Language (formerly called Flemish-Belgian Sign Language) started in the 1990s, the demand for a bilingual dictionary grew. From 1999 onwards, a number of small-scale lexicographic projects were set up all over Flanders, which resulted in the publication of the first online Dutch – Flemish Sign Language (VGT) dictionary [gebaren.ugent.be](http://gebaren.ugent.be) in 2004.

In this bilingual dictionary, each sign was represented in three ways: a video clip, a Dutch translation, and Signwriting (a notation system that allows writing down signs in a visual way) (Van Herreweghe, 2001). This allowed for bidirectional search options, meaning that users could search for a sign by selecting its handshape, location or movement (based on the Signwriting), by scrolling through the Dutch words in alphabetical order, or by typing a Dutch word into a search bar (Vermeerbergen and Van Herreweghe, 2018). This was the first generation sign language dictionary in Flanders. For more information on the aims and the methodology of this dictionary, please see Vermeerbergen and Van Herreweghe (2018) and De Weerd et al. (2003).

From 2012 until the launch of the new dictionary in 2019, [gebaren.ugent.be](http://gebaren.ugent.be) was managed by the Flemish Sign Language Center (Vlaams GebarentaalCentrum / VGTC), the center of expertise for Flemish Sign Language. VGTC has pursued a further lexical extension

of the dictionary since then, adopting the guidelines explained in the lexicographical research methodology of Oyserman et al. (2012).

VGTC has been lobbying for a thorough revision of the dictionary’s interface. It seemed that, almost 15 years after the launch of the first dictionary, users’ expectations were not fully met. After all, electronic technology had evolved very rapidly since 2004, greatly expanding the possibilities for an accessible online reference tool (McKee and McKee, 2013). Mixed methodology user research supported this hypothesis, as is explained in more depth later.

At the end of 2018, VGTC was awarded project funding, which enabled its employees, in collaboration with a software development company, to implement the large-scale renewal of the dictionary. In doing so, a practical lexicographical perspective was adopted: describing the language in a way that is faithful to the available lexicographic research, and always taking into account the expectations, needs and skills of those who will use the dictionary (Atkins and Rundell, 2008).

## 2. Developing the New Dictionary

### 2.1. User Research

When VGTC received the one-off project funding for the renewal of the online dictionary in 2018, the first meaningful steps were taken towards a new interface. In particular, extensive research regarding the users of the dictionary was conducted. More than wanting



Figure 1: A screenshot of the home page of gebaren.ugent.be

to project a report of lexicographic research on a website, the goal was to make the new dictionary a practical, user-friendly reference tool that meets the needs, expectations and skills of the dictionary users. To gain a better understanding of who the users were, several sources were consulted: the user research by Oyserman (2013), the quantitative data from Google Analytics and VGTC's own user profiles.

An initial qualitative user study of gebaren.ugent.be was conducted by Joni Oyserman in 2013. In this preliminary study, a group of deaf and hearing teachers of VGT, interpreters, students of VGT, deaf and hearing parents of a deaf child and relatives of a deaf person were questioned about their usage pattern of the dictionary. One of the results of this survey is that users like to search from VGT to Dutch as well as vice versa. They would also like to search thematically to see all semantically related signs. Since the latter was not yet possible, this was taken into account and it was added during the development of the new VGT dictionary.

In addition, data from Google Analytics, a service from Google that gives website administrators access to usage statistics for that application, were also analysed. This tool keeps track of which signs are viewed the most, how long users visit a certain page and how they search. The data analysis showed that the search function VGT to Dutch was used only for a small percentage of the searches. This seems to be diametrically opposed to the result from the qualitative user study described above, in which informants indicated that this search direction is important. A possible explanation might be that the SignWriting symbols are not sufficiently accessible to the average user. Another assumption is that the search function yields too extensive a set of results to quickly arrive at the desired entry. In an attempt to make the VGT to Dutch search function more user friendly, it was decided SignWriting would no longer be used in the new dictionary.

Moreover, the percentage of new visitors was considerably larger than the number of returning visitors,

which could indicate that the dictionary's content was, to some extent, inadequate. Also, whereas the website's interface was not responsive, meaning the web application is less accessible on smartphones and tablets, dictionary visitors did tend to use a mobile device in almost half of the sessions. Therefore, the need for a responsive website, which automatically adapts to other screen sizes and thus remains clear and user-friendly on different devices, became apparent. Because of this, it was necessary to develop a new web application.

In collaboration with AE, the software company that developed this new web application, fictitious user profiles were created to test the hypothesis from the above studies. During the development process, each version of the application was also tested by a small group of users. In April 2019 a testing session was held, in which eight volunteer informants (two deaf elderly, two young deaf people, two interpreters and two hearing interpreting students) carried out various assignments for one hour on the first version of the application. Their findings were included in the further development of the dictionary interface to make it as functional as possible. For instance, the testing session revealed the first iteration of the new VGT to Dutch search function needed to be adjusted.

## 2.2. Data Cleaning and Modelling

Since 2017, VGTC has been using Signbank, an electronic database specifically developed to compile and manage lexicographic data for sign languages. The precursor to Signbank originated in Australia by the work of Johnston (2001), and then further developed for AUSLAN and other sign languages like BSL (Cormier et al., 2012) and NGT (Crasborn et al., 2016). As sign languages are visual languages, it is important to use a database which allows for easy uploading and processing of video clips. In addition, this system offers the possibility to annotate signs morphologically, phonologically and semantically. These last two were used extensively, for the new VGT to Dutch search option and the thematic categories respectively.

Signs are added to Signbank by team members from various sources. For each sign that is added, the source is specified to enable further specific research. All data found in *gebaren.ugent.be* were added to Signbank, as were the results of the lexicographical research projects conducted by VGTC after 2004. Furthermore, the Signbank database enables VGTC to more easily collect signs on a larger scale through community sourcing. For example, employees can follow up on discussions about certain signs in Facebook groups and copy the data into Signbank. It should, however, be noted that this way of community sourcing is time consuming because of the manual effort it requires. In any case, Signbank has become the central hub where all lexicographical data for VGT are collected. These data are now used to directly feed the new dictionary. Currently Signbank holds about 20,000 entries, about half of which are published in the online dictionary.

Each entry contains a unique gloss, a gloss in this case being the closest representation of the meaning of the sign in written Dutch. One gloss represents one concept. The signs for both “poor” and “(an) arm” could be glossed as ARM in Dutch so in order to avoid confusion, “poor” was glossed as ARMOEDE, i.e. “poverty”, a different possible translation of the same sign, instead. One gloss/concept can have several sign variants, in order to keep the glosses unique, a serial letter is added after each gloss, e.g. ARM-A, ARM-B, ... These glosses are only used in Signbank. The Dutch to VGT search bar on the dictionary uses the possible translations added to each entry.

Bringing together all these raw data from different sources inadvertently led to inconsistencies and small mistakes, for example: duplicate signs (i.e. signs which were accidentally added twice under a different gloss), inconsistent or confusing glosses (e.g. two concepts, like “poor” and “arm” under the same gloss), missing information, typos, etc. Because of this, the data had to be manually revised and complemented. Furthermore, in 2018 and 2019, part of the old video recordings of signs were renewed in Signbank to increase the video quality. This work will be continued until all recordings are dealt with.

Because SignWriting was not deemed very transparent, efficient and widely accessible, it would no longer be included in the new dictionary. In view of the new search function from VGT to Dutch, a phonological annotation was performed based on the hand shape and location of the signs. The sets of 34 handshapes and 20 locations were selected based on the research of Demey (2005). While the available SignWriting images were originally to be used for this annotation, it turned out this was not practically or technically feasible. In other words, all entries were annotated manually, mostly by the researching team (two deaf

and one hearing) and to a lesser extent a volunteer (hard-of-hearing).

In order to enable thematic searches, signs were also assigned one or more semantic categories (e.g. nature, law, sports, medicine, family, ...) during this phase. The original lexicographic approach which was used to compile the previous dictionary, is at the basis of the semantic categories in the current dictionary. For more information, please see Vermeerbergen and Van Herreweghe (2018). The original list of semantic categories was reduced and reworked. There were three possibilities: (i) the category would be kept (and maybe renamed), (ii) the category would be merged with another category, or (iii) the category would cease to exist.

To decide which signs from Signbank are added to the dictionary, a committee of carefully selected deaf near-native signers (two from each of the five provinces in Flanders) meets 3 to 4 times a year. At least one of VGTC’s employees is also present to guide and moderate the discussion. At these meetings, Signbank entries tagged as “*expertgroep*” (“group of experts”) are discussed. The rather small size of the group allows for thorough and in-depth discussions. Leading up to the discussion, all participants received the necessary information to prepare in advance (i.e. think about the signs and concepts to be discussed and check with their own network of signers). Each participant would give a final “yes” or “no” at the end of the discussion as to whether to include the sign in the dictionary or not. If the sign was confirmed to be present in only some of the provinces, it would be included as a regional variant. However, the qualitative nature of this evaluative entity also means it is rather time consuming and inefficient, due to the practical difficulties of bringing everyone together physically and the manual processing of the data. It is therefore preferable to supplement it with quantitative data from the VGT corpus.

“In the last decade, much care, time, and resources have been invested in compiling the VGT corpus. This corpus consists of 5TB or 140 hours of video data produced by 120 deaf L1 signers” (Brosens et al., 2021). Approximately 40 hours of data was transcribed so far (Wille et al., 2022). Inspired by Crasborn et al. (2016), VGTC is still working on a link between Signbank and ELAN, the annotation program used by the annotators of the corpus. Through this form of corpus linguistics, the aim is to strengthen the lexicographic basis of Signbank in the future. Currently the corpus is used as much as possible for lemma selection (and thus to a lesser extent confirmation of existence). When – while annotating the corpus – signs are found that were not yet collected in Signbank, the annotator adds this sign manually. Future uses of the corpus hopefully include:

example material, sense discrimination, and description of regional variation. In the meantime, alternative ways of confirming the spread / existence of signs are being explored, for instance online questionnaires or polls using existing platforms or alternatively building a custom platform for this very purpose.

### 2.3. Innovations

As shown in figure two, the VGT dictionary was mainly formally modernised. However, there are also a number of substantive differences with regard to the previous dictionary. First of all, the search options were expanded. Because of this, users of the dictionary can now arrive at a certain sign in different ways. Just like in *gebaren.ugent.be*, people can search the Dutch equivalent of a sign in VGT via the hand shape or location of the sign. However, this no longer involves SignWriting, but images of the hand shapes and the locations on the body. Accessible information icons explain in a comprehensible way, both in VGT and in Dutch, what hand shapes and locations exactly are. Users can search Dutch to VGT via a search bar, in which a Dutch word can be typed. In addition, there is also the possibility to search by regional variant or by semantic category. Finally, a combination of search filters is possible with the aim of making the users of the dictionary find the searched entry faster.

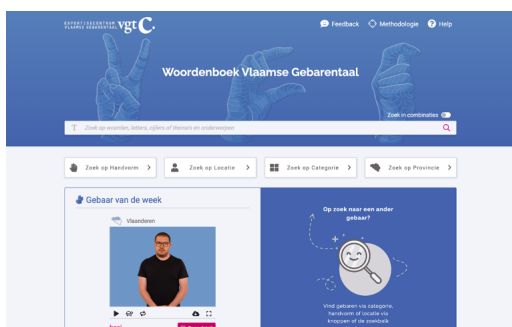


Figure 2: A screenshot of the home page of the new dictionary

Once the searched entry is found, one click reveals the detail page where the phonology and semantic category of the sign are displayed. Thanks to cross-referencing, users are easily directed to signs that have the same meaning, but are used in a different region, or to signs that are phonologically related.

An interesting extension in the new application is the possibility to link different Dutch words to one lemma. In *gebaren.ugent.be* synonyms in Dutch (e.g. “climbing” and “scrambling”) were placed under two different lemmata. The same applies to different parts of speech in Dutch (e.g. “relax” and “relaxation”). In the new dictionary, multiple possible Dutch translations can now be linked to a single sign.

As administrator of the new dictionary, VGTC is also more in touch with the user. Users can provide feedback, ranging from technical problems (e.g. a video that does not play properly) to more substantive comments (e.g. I suspect this sign is also used in my region), in an accessible way, through video and/or text. The sign is then appropriately tagged in Signbank and further research can be conducted. In addition, VGTC also receives a notification if users type a word in the dictionary and find no result. That way VGTC gets an overview of which lemmata users find lacking in the dictionary.

In addition to these substantive changes, the website has also been structurally changed. As mentioned earlier, the website is now responsive, so the interface adapts to the device on which people visit the website. In this way, the dictionary is as accessible on a smartphone or tablet as it is on a computer screen. The entire interface is also more visually organised. This is mainly the result of an initial user test. It was found desirable to avoid the Dutch text becoming a barrier for part of the target group.

### 3. Conclusions

Evidently, making a user-friendly bilingual and bidirectional online dictionary is a never ending process. VGTC aims to continuously improve both the user based interface and the content of the current dictionary. Future goals include, but are not limited to, adding definitions and sample sentences (preferably extracted from the corpus), as well as information on the etymology and common use of signs.

Even if the development of this new dictionary is a step forward, there are also limitations. It cannot be guaranteed, just like in *gebaren.ugent.be*, that every existing sign in VGT is included in the dictionary. Therefore, signs used by a sign language user that are not in the dictionary are no worse than or inferior to the ones found in the dictionary. Just as in the UGent project, VGTC naturally strives for a description of the language that is as complete as possible and is constantly working on expanding and deepening the dictionary.

As Atkins and Rundell (2008) state, “the content and design of every aspect of a dictionary must, centrally, take account of who the users will be and what they will use the dictionary for”. VGTC, too, strongly believes that more in-depth users research, preceding, during and after the development of a sign language dictionary, is crucial in order to build a sustainable reference work, which dictionary users can continue to explore and enjoy using.

#### 4. Bibliographical References

- Atkins, B. and Rundell, M. (2008). *The Oxford Guide to a Practical Lexicography*. Oxford University Press Inc., New York.
- Brosens, C., De Durpel, H., Beukeleers, I., and Wille, B. (2021). Towards a functional label set for the on-line dictionary of flemish sign language. In *FEAST*, pages 50–60, Hong Kong.
- Cormier, K., Fenlon, J., Johnston, T., Rentelis, R., Schembri, A., Rowley, K., Adam, R., and Woll, B. (2012). From corpus to lexical database to online dictionary: Issues in annotation of the bsl corpus and the development of bsl signbank. In Crasborn, O., et al., editors, *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon (LREC'12)*, pages 7–12, Istanbul, Turkey. European Language Resource Association (ELRA).
- Crasborn, O., Bank, R., Zwitserlood, I., Van der Kooij, E., Schüller, A., Ormel, E., Nauta, E., van Zuilen, M., van Winsum, F., and Ros, J. (2016). Linking lexical and corpus data for sign languages: Ngt signbank and the corpus ngt. In Efthimiou, E., et al., editors, *Proceedings of the 7th Workshop on the representation and processing of sign languages: Corpus Mining (LREC'16)*, pages 41–46, Portoroz, Slovenia. European Language Resource Association (ELRA).
- De Weerd, K., Vanhecke, E., Van Herreweghe, M., and Vermeerbergen, M. (2003). *De dovengemeenschap in Vlaanderen : Doorlichting, sensibilisering en standaardisering van de Vlaamse gebarentaal : Eindrapport, luik 2. Op onderzoek naar de Vlaamse gebarentaal*. Fevlado-Diversus, Gent, BE.
- Demey, E. (2005). *Fonologie van de Vlaamse Gebarentaal: distinctiviteit en iconiciteit*. Phd thesis, Universiteit Gent, Faculteit Letteren en Wijsbegeerte, Gent.
- Johnston, T. (2001). The lexical database of auslan (australian sign language). *Sign Language and Linguistics*, 4(1-2):145–169.
- McKee, R. and McKee, D. (2013). Making an online dictionary of new zealand sign language. *Lexikos*, 23(1):500–531.
- Oyserman, J., Heyerick, I., and Huys, E. (2012). *Onderzoeksmethodologie: Datacollectie voor lexicografisch onderzoek van de vlaamse gebarentaal*. Vlaams GebarentaalCentrum.
- Oyserman, J. (2013). *Enquête resultaten gebruikersonderzoek online woordenboek vlaamse gebarentaal*. Vlaams GebarentaalCentrum vzw.
- Van Herreweghe, M. (2001). *SignWriting in de Vlaamse Gebarentaal*. Cultuur voor Doven, Gent, BE.
- Vermeerbergen, M. and Van Herreweghe, M. (2018). Looking back while moving forward: The impact of societal and technological developments on flemish sign language lexicographic practices. *International Journal of Lexicography*, 31(2):167–195.
- Wille, B., Beukeleers, I., Van Herreweghe, M., and Vermeerbergen, M. (2022). Big things often have small beginnings: on the development and use of small and big corpora for flemish sign language linguistic research. *Frontiers In Psychology*, 12. DOI: 10.3389/fpsyg.2021.779479.