

# Hand in Hand – Using Data from an Online Survey System to Support Lexicographic Work

Sabrina Wähl, Gabriele Langer, Anke Müller

University of Hamburg, Institute for German Sign Language and Communication of the Deaf, Germany





- first used to verify signs and their presumed meanings in previously published sign collections (Langer et al., 2014)
- currently primarily used to supplement corpus data

### Sampling of the DGS-Feedback

- the system is open to all members of the DGS community
- up to now 279 persons contributed to the DGS-Feedback (February 2018)
- registration to obtain metadata on person and sign language skills
- heterogenous group of participants e.g. deaf, hard of hearing, CODAs, interpreters, early learners and late learners

#### **Structure of the Survey**

- different question types
- type 1:
- asks for different meanings of one sign form
- type 2:
- asks for different signs used for the same concept
- a questionnaire (work package) consists of several question pages (questions) and a question consists of several question items

# Type 1

- first question type released
- 42 work packages released

	Ξ,
unknown	4

• strong corpus evidence would already lead to inclusion of the sense into the dictionary • high positive DGS-Feedback response also confirms the finding

# Weak Corpus Evidence and Low Positive DGS-Feedback Response:

Example 2					
'Monday' ✑ ┍ ₀ ◡ <sup>〔 X</sup> ¹ ǜ Ì ÷ +					
sense	Monday, name of the first day of the week				
number of corpus tokens	9 (from 4 informants)				
total number of responses	104				
used	5				
known	15				
unknown	84				



#### • contemplating the results separately both may not be conclusive

unknown	54	



discrepancy between corpus findings and DGS-Feedback data

• may be result of the different sampling of corpus and DGS-Feedback

• relatively high corpus token count results from an elicitation task (asking for certain concepts that were assumed to be highly variable), still 7 tokens for '(loaf of) bread' occur in tasks with conversational character

• corpus data suggests that the sign-sense combination is mainly used in Bavaria and Hesse • DGS-Feedback data confirms the use in Southern Bavaria; for Hesse data is still scarce (the only participant from that area voted unknown)

• "core area" of distribution appears to be Southern Bavaria and an according note in the dictionary will be made

# Weak Corpus Evidence and

High Positive DGS-Feedback Response

Example 5

'earring' ∋<sub>~02</sub> (X<sub>12</sub> 8<sub>√</sub>)

- → over 100 returns: 14 with 71 different sign forms
- most returns from type 1 work packages (first presented to new users)
- stimuli: a clip of a single sign, written German equivalents and if necessary a signed context
- three response options: I use the sign (used), I know the sign (*known*), I do not know the sign (*unknown*)
- at the end of each question users may bring other meanings to our attention (by writing or webcam recording)





Stimulus and answer-buttons for one form-meaning combination in the DGS-Feedback

Structure of questions

# **Analysis Stage of Corpus- based** Lexicographic Work

- with an increasing number of tokens available analysis stage of dictionary making started
- ➡ data analysis and documentation of relevant facts about the sign
- → lemmatisation (lemma establishment; Svensén, 2009)
- ➡ description of a signs' meanings and grouping them into senses (word sense disambiguation -WSD; cf. Atkins & Rundell, 2008:269)

- considering them together with regard to region, age, hearing status, and age of language acquisition can lead to stabilisation of findings and thus suggest an answer
- in case of 'Monday' the low proportion of corpus findings and used answers appear to be the result of a very regional distribution in Lower Saxony
- although 330 informants were recorded, information on regional signs may be scarce
- DGS-Feedback participants add with their information on use and knowledge
- because of token numbers the sense ,Monday' would not yet be included into the dictionary

### **No Corpus Evidence and**

# No or Low DGS-Feedback Response

E	Example 3
'menu' ◯ <sup>2</sup> r 0 ⇔ <sup>↓</sup> )(+	
sense	small display on the computer to choose editing options
number of corpus tokens	0
total number of responses	103
used	3
known	14
unknown	86

• core meanings of the sign  $\bigcirc^{\frac{1}{2}}$ ,  $0 \bigcirc^{\frac{1}{2}}$ ,  $1 \odot^{\frac{1}{2}}$ ,  $1 \odot^{\frac{1}{2}$ • another meaning is ,menu' (sense: list or range food offered)

Example 7

sign:  $\exists r \circ \infty \bullet \setminus j \bullet^{j(\pm \chi + \chi)}$ 

core sense

- spoken German Menü (basis for the corresponding mouthing) can also denote menu options of a computer program, leading to the possible assumption that  $\bigcirc^2 r \circ \bigcirc^{\frac{1}{2}}$  also could be used for 'menu (computer)' - as has been published in a sign collection prior to the Korpus-Project • the sense 'menu (computer)' was included in the DGS-Feedback to check on the use of this sense
- results for 'menu (computer)' show, that this sense-sign combination is very likely not established

sense	jewellery worn on the ear
number of corpus tokens	6 (from 4 informants)
total number of responses	139
used	121
known	14
unknown	4



 sign-sense possibly a low frequency item, or no relevant topic came up during elicitation • DGS-Feedback data provides us with a good reason to include the sense (otherwise it would be held back until token count increases)

## **No Corpus Evidence and**

# High Positive DGS-Feedback Response

Example 6			
'medical' [∂,] )(↓ X +			
sense	of a or concerning a doctor		
number of corpus tokens	0		
total number of responses	124		
used	87		
known	19		
unknown	18		

• this case leads to a preliminary description of the sense within the pre-dictionary database (status *under surveillance*) but not to inclusion into the dictionary

• inclusion into the dictionary will follow in case corpus evidence emerges as we work on a corpusbased dictionary and example sentences are taken directly from the corpus

# **Participants' Comments on Sign Use**



• through comments interesting insights in homonyms and form-related signs e.g. 'eye' and 'to try'

• information enriches the lexical database (cross-references between types are made) and helps annotation

- description of form variants and regional distribution of signs (McKee & McKee, 2013; Zwitserlood et al., 2013; Fenlon et al. 2015)
- corpus data is our starting point
- corpus data has priority over DGS-Feedback data
- corpus data can only provide positive evidence and areas of uncertainty remain
- data from the DGS-Feedback supplement corpus findings and thus support lexicographic work

form-related sign: $\exists F 0 \exists F \rangle = \mathcal{I}( \pm \chi + 1)$		A second s
core sense	'to try'	
number of written	3	
		()

'to try' 'eye´

• the dictionary will include cross-references to form-related signs video comment is used if participants do not agree with a sign-meaning combination being presented and thus give their sign for the meaning video comments can be spot transcribed and hence supplement corpus data

# Conclusion

- The DGS-Feedback data is a valuable addition to the corpus findings.
- The benefits of findings are: confirmation of uncertain sign use and showing certain characteristics of a sign (e.g. regional use, form variation, age effects).
- Question types 1 and 2 were designed to verify or disprove non-corpus data. With corpus-based lexicographic work there is a need for new question types.
- When analysing data from both sources several factors need to be considered as we have shown in examples 2, 4 and 6.

#### References

Atkins, B.T.S., Rundell, M. (2008). The Oxford Guide to Practical Lexicography. Oxford: Oxford University Press. Fenlon, J., Cormier, K., Schembri, A. (2015). Building BSL SignBank: The Lemma Dilemma Revisited. International Journal of Lexicography, Volume 28, Issue 2. Oxford: Oxford University Press, pp. 169--206.

Langer, G., König, S., Matthes, S. (2014). Compiling a Basic Vocabulary for German Sign Language (DGS) - lexicographic issues with a focus on word senses. In A.Abel, C. Vettori, & N. Ralli (Eds.), Proceedings of the XVI EURALEX International Congress: The User in Focus, July 15-19 2014 in Bolzano/Bozen - Italy. Bolzano/Bozen: EURAC research, pp. 767--786.

McKee, R., D. McKee. (2013). Making an Online Dictionary of New Zealand Sign Language. Lexikos, 23.1. Stellenbosch: Buro van die Wat, pp. 500--531.

Svensén, B. (2009). A Handbook of Lexicography. The Theory and Practice of Dictionary-Making. Cambridge: Cambridge University Press.

Zwitserlood, I., Kristoffersen, J., Troelsgård, T. (2013). Issues in Sign Language Lexicography. In Jackson, H. (Ed.), The Bloomsbury Companion to Lexicography. London: Bloomsbury Publishing, pp. 259--283.

**Poster presented** at the 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. LREC 2018, Miyazaki, Japan. May 12th, 2018.

This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.