Exploring Lexical Variation in a Growing Corpus of DGS

Sabrina Wähl, Gabriele Langer, Anke Müller, Julian Bleicken, Thomas Hanke, Reiner Konrad

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

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

AKADEMIE DER WISSENSCHAFTEN IN HAMBURG

DGS-KORPUS

DGS Corpus recorded 2010-2012
- Number of informants 330
- Controlled sample balanced for
  - gender
  - 13 regions
  - 4 age groups: 18-30, 31-45, 46-60, 61+
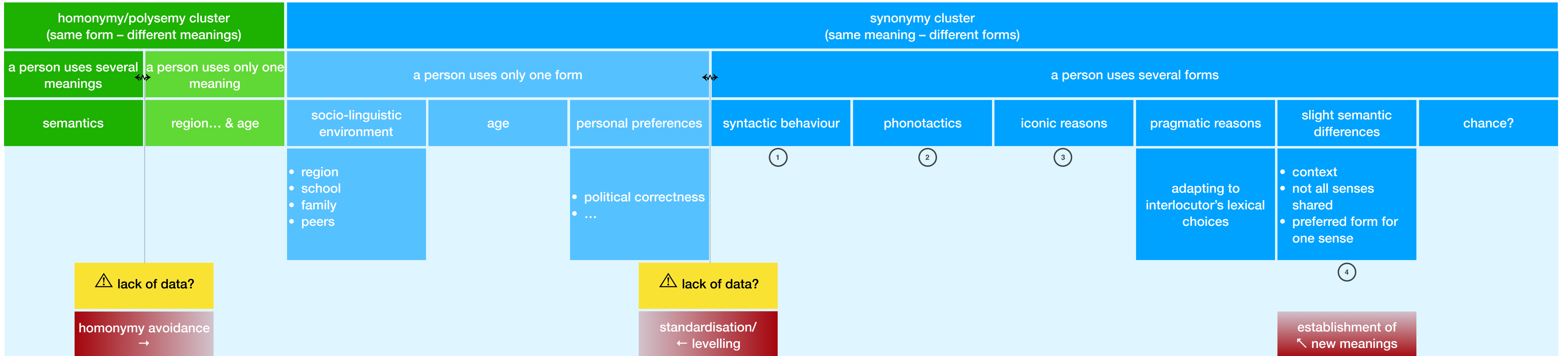- Native and near-native signers

Data
- Filmed conversations and staged communicative events (Nishio et al. 2010)
- Multi-modal corpus, lemmatised and accessible through iLex (Hanke/ Storz 2008)
- About 560 h footage of natural signing
- Lemmatised: 576.400 tokens (2019-09-23)

Starting Point
- The size of our corpus supports analyses of regional variation. Regional distribution of lexical variants of roughly synonymic sign clusters can easily be visualised on maps (cf. Hanke et. al. 2017).
- However: Often several competing signs of a sign cluster are used within the same region and even by a single individual.

Question
- Looking beyond regional and sociolinguistic background: What other factors influence the lexical choice of signers?

homonymy/polysemy cluster (same form – different meanings)		synonymy cluster (same meaning – different forms)								
a person uses several meanings	a person uses only one meaning	a person uses only one form			a person uses several forms					
semantics	region… & age	socio-linguistic environment	age	personal preferences	syntactic behaviour ①	phonotactics ②	iconic reasons ③	pragmatic reasons	slight semantic differences	chance?
		• region • school • family • peers		• political correctness • …				adapting to interlocutor's lexical choices	• context • not all senses shared • preferred form for one sense ④	

⚠ lack of data? → homonymy avoidance →

⚠ lack of data? → standardisation/ ← levelling

establishment of ↖ new meanings

Lexical Choice - Homonymy Avoidance

Regional distribution of lexical variants found e.g. for signs for 'girl' seems to corroborate the hypothesis of homonymy avoidance (see Gilliéron & Roques 1912, and for signed languages Boyes Braem 1981, Cuxac 2000).

But:

Corpus data also show lemma pairs or clusters of homonymous signs in the same region.

'Friday' — 17 informants
'girl' — 127 informants
'girl' — 73 informants

'class' — 36 informants
'wood' — 23 informants
'why' — 14 informants

'woman' (one of 10 lexical variants) compared to 'bread' (one of 22 lexical variants)

'woman':
- age group 61+, 5 informants
- age group 46+, 23 informants
- age group 31+, 34 informants
- age group 18+, 49 informants

'bread':
- age group 18+, 19 informants

Apparent time map (cf. Hanke et al. 2017) of one variant sign for 'woman' also may suggest that homonymy avoidance plays a role in regional language change. In Bavaria and Hesse the distribution of this meaning seems to be blocked by a homonymous sign for 'bread'.

Political Correctness / Age Variation

AFRICA1

In the case of 'Africa' the preferred use of a lexical variant AFRICA1 (used by 21 informants) in comparison to AFRICA2 (used by 4 informants) is attested.

AFRICA2

This is a case of age variation. AFRICA2 apparently is becoming obsolete. This may be due to the fact that it is perceived as politically incorrect.

Use of signs for 'Africa' by age groups: 25 informants (with 43 tokens)

AFRICA1 | AFRICA2 | all: AFRICA

- 18-30
- 31-45
- 46-60
- 61+

Lexical Choice - within Semantic Clusters

Signer X from the Hamburg area, age group: 60+, 35 tokens of 4 forms

Synonymy cluster for 'speak, talk, say, language'
- 5 different signs located at mouth (similar iconic motivation)
- Includes 2257 tokens from 293 persons in the corpus
- Overlapping meanings
- No clear regional distribution: several signs used in each region
- Assumption: slight meaning differences
  - investigation: closer look at use by one person "signer X" reveals use of different forms for different meaning aspects for this one person

	TO-SAY1		TO-SPEAK1		LANGUAGE1		LANGUAGE2		LANGUAGE3	
	corpus	signer X	corpus	signer X	corpus	signer X	corpus	signer X	corpus	signer X
tokens	1529	17	372	5	296	6	191	7	169	0
mouthings include forms of	sagen	sagen	sprechen, sagen	sprechen	sprache, sprechen, sagen	sprache, sprechen	sprechen, sprache	only mouth gesture	sprechen	
predominant meaning as used by signer X	independent of hearing status or language used: introducing the content of an utterance, citation or opinion of somebody, e.g. 'she said …'; focus on content		hearing person speaking in a spoken language to Deaf person(s), some using especially articulated lip movements or supporting their speaking by gestures; focus on manner of (visible) articulation		reference to specific (spoken) language(s); Deaf person speaking a spoken language; focus on ability to speak (also used as element of loan compounds)		hearing person(s) use spoken language while Deaf person(s) present do not have access to the content (usually in group situations such as school or in mixed groups); focus on inaccessibility of content		not used by X	

Synonymy cluster: TOGETHER

lexical variants

	TOGETHER1A + 2 variants	TOGETHER3A	TOGETHER6	TOGETHER-PERSON1
token count	552	230	39	126
semantic difference	together: in a group ④	together: two persons	together: two persons/ two parties (abstract)	together: two persons
polysemy	'group' (328 corpus tokens), 'community' (43 corpus tokens)	'with' (1075 corpus tokens)	'with' (31 corpus tokens), sign becoming obsolete	morphologically related signs: TO-ACCOMPANY1A, TO-SEPARATE4B
syntactic behaviour	spatial modification (1 locus)	deictic use, spatial modification (2 loci)	spatial modification (1 locus)	spatial modification (1 or 2 loci)
iconicity	depicting handshape: size & shape	no depicting handshape		depicting handshapes: '2 persons'

- Contrastive analysis of meaning cluster
- Individuals using several items of a cluster tend to distinguish between different senses and functions

Strategy of individual informants using both:
- difference in semantic roles indicated (person vs. place) ①
- style: two-handed form allows for enlarging
- visual context: build coherent units using same handshape (handshape harmony) ②

Strategy:
- use in context of two specific person referents, add depicting constructions in context ③

Conclusion:
- Looking into individual language use can help to find usage distinctions in semantic clusters. Results need to be validated against a larger number of corpus informants: Across informants, more meaning overlaps can be observed.
- Apparent time only allows a rather coarse diachronic view on the data, competing processes like establishment of new meanings and levelling would need a finer granularity on the timeline to be separated. Exact synonyms (lexical variants) are rare, if not regionally distributed.
- Homonymy avoidance cannot be claimed as a general rule, but we find data fitting the pattern.

References:
Boyes Braem, P. (1981): Features of the Handshape in American Sign Language. Ph.D. thesis, University of California, Berkeley.
Cuxac, C. (2000): La langue des Signes Française (LSF). Les voies de l'iconicité. Paris: Ophrys.
Gilliéron, J. / Roques, M. (1912): Etudes de géographie linguistique d'après l'Atlas linguistique de la France. Paris: Champion.
Hanke, T. / Storz, J. (2008): iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, & E. Thoutenhoofd (Eds.), Construction and Exploitation of Sign Language Corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco. Paris: ELRA, pp. 64-67.
Hanke, T. / Konrad, R. / Langer, G. / Müller, A. / Wähl, S. (2017): Detecting Regional and Age Variation in a Growing Corpus of DGS. Poster presented at the workshop "Corpus-based approaches to sign language linguistics: Into the second decade", Birmingham UK, 24 July, 2017.
Nishio, R. / Hong, S. / König, S./ Konrad, R. / Langer, G. / Hanke, T. / Rathmann, C. (2010): Elicitation methods in the DGS (German Sign Language) Corpus Project. Poster presented at the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, following the 2010 LREC Conference in Malta, May 22-23, 2010. Workshop Proceedings. W13. 4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. May 22/23, 2010. Valetta – Malta. Paris: ELRA. pp. 178-185.