DGS Corpus Project
Development of a Corpus Based Electronic Dictionary
German Sign Language / German

Thomas Hanke, Susanne König, Reiner Konrad, Gabriele Langer, Christian Rathmann

State of the art:
• no large corpus of DGS available
• no comprehensive dictionary of DGS on linguistic principles available
• many sign collections of DGS available, differing in size, standard and degree of documentation
• several smaller project-based corpora from technical sign dictionaries available
• complex lexical database system (Lex) available
• experience and know-how from more than 10 years of dictionary making in Hamburg
• several other national corpus projects under way in or preparation
  • cooperations, transfer/exchange of know-how, workshops

DGS Corpus
• informants from all over Germany of different age, sex and social status;
• selected on grounds of language competence and regional rootedness;
• comparable to large spoken language corpora in size & scope;
• corpus should reflect a representative and well-balanced part of everyday communication;
• design allows the use of the corpus for various tasks:
  • validation of the basic vocabulary;
  • thorough research on DGS grammar based on the transcriptions;
  • identification of different meanings and collocations of a sign;
  • the design anticipates a comparative sociolinguistic study comparable in kind and quality to Lucas et al. (2001) and Schiembr/Johnston (2004).

Focus group
• direct representatives of the language community;
• approx. 20 deaf experts from different regions;
• trained and sensitized to linguistic questions;
• consulted in cases of doubt concerning, e.g.: the lexical status of a sign, citation form, language change, sign use and (regional) variation;
• decisions concerning lemma selection or the well-formedness of grammatical constructions continuously validated by focus group;
• filling gaps, avoiding additional surveys;
• approximately 6000 lemma entries expected

Analysis and compilation of dictionary entries
• abstracting from occurrences and all gathered data on a sign to obtain a description of lexical items;
• description of regular grammatical properties of clauses of lexical items;
• description of form-phonological variants, use of space, morphology, word class/syntactic functions, meaning);
• possible translations into German, iconic value & visualisation technique, dialectal information, cross-references to related and similar signs, synonyms, antonyms, examples of use

Analysis
• detailed transcription of surrounding context of a sign, instances of grammatical uses of a sign, different sign forms
• segmentation into utterances
• transcription of surrounding context
• translation

Basic transcription
• transcription into German
• segmentation into utterances
• contents accessible via German

Detailed transcription
• transcription into English
• segmentation into utterances
• contents accessible via German
• tagging of phrases and sign strings (within utterances)
• coding and annotation of productive sign forms
• more detail: transcription of surrounding context of a sign

Comprehensive, corpus based, electronic dictionary DGS-German
• contains approx. 6000 sign entries
• all contained information is corpus based
• entries contain descriptive information on signs and uses taken from the corpus
• dictionary focus on DGS part
• German part provides access to DGS for hearing users;
• deaf users can get basic information about German words
• combination of different types of dictionary functions
• primarily serves the following target groups:
  • DGS learners, e.g. hearing parents of deaf children, students of Deaf Studies/signal language interpreting
  • professional sign language interpreters for DGS – German
  • native signers of DGS, deaf adults, CODAs
  • deaf children or pupils, acquiring DGS as native language
  • sign language teachers, linguists and others concerned with sign language structure

Dictionary grammar
• based on the corpus and information from the focus group
• general overview of the most important grammatical features of DGS, supplemented by examples taken from the corpus
• written in easily understandable terms
• sign entries become shorter, more compact and therefore more clearly

Preliminary basic vocabulary
• basic vocabulary of DGS and German
• not based on corpus data but on published sign collections
• signs verified by focus group and public voting
• to be replaced by the general, comprehensive dictionary after 10 years

Published sign collections
• no documented empirical basis
• valuable source for basic vocabulary
• a pool of signs to be evaluated by linguists and native signers
• compilation of a preliminary basic vocabulary
• for DGS to be published after 1 years
• ongoing validation by corpus data

Published collections of signs
• approx. 350-400 hours
• 250-300 informants
• approx. 2.25 m. tokens

Data collection
• 12 different locations (see map): areas of relatively high deaf population density (deaf schools, deaf centres), which are easy to reach from surrounding rural areas;
• mobile studio is successively set up in each of the locations;
• each elicitation: pair of two informants
• peer-to-peer procedure
• interviews conducted by a local deaf contact person to secure an elicitation of regional sign variants with as little influence from the interviewer as possible;
• each elicitation consists of:
  • a standardized interview covering language and social data (approx. 20 min. each); aim: description of data on a metadata level – IMDI standard
  • filming of spontaneous conversations on a given topic (2-3 topics from a list, 60-90 min. for each pair of informants);
  • elicitation of as much basic vocabulary as possible; list of topics covers about 20 subject areas.
• different tasks with selected stimuli (approx. 120 min. for each pair).

Tasks aim at capturing special phonological, morpho-
• logial, syntactic and lexical phenomena.
• setting: see illustrations
• cameras 1-5: HD 1080
• camera 6-7: stereoscopic 480p48

Basic transcription
• translation into German
• segmentation into utterances
• contents accessible via German

Detailed transcription
• translation into English
• segmentation into utterances
• contents accessible via German

Publication of the first comprehensive, corpus
• based, electronic dictionary DGS-German

Poster presented at the Corpus Linguistics Conference in Liverpool, 20 – 23 July 2009