

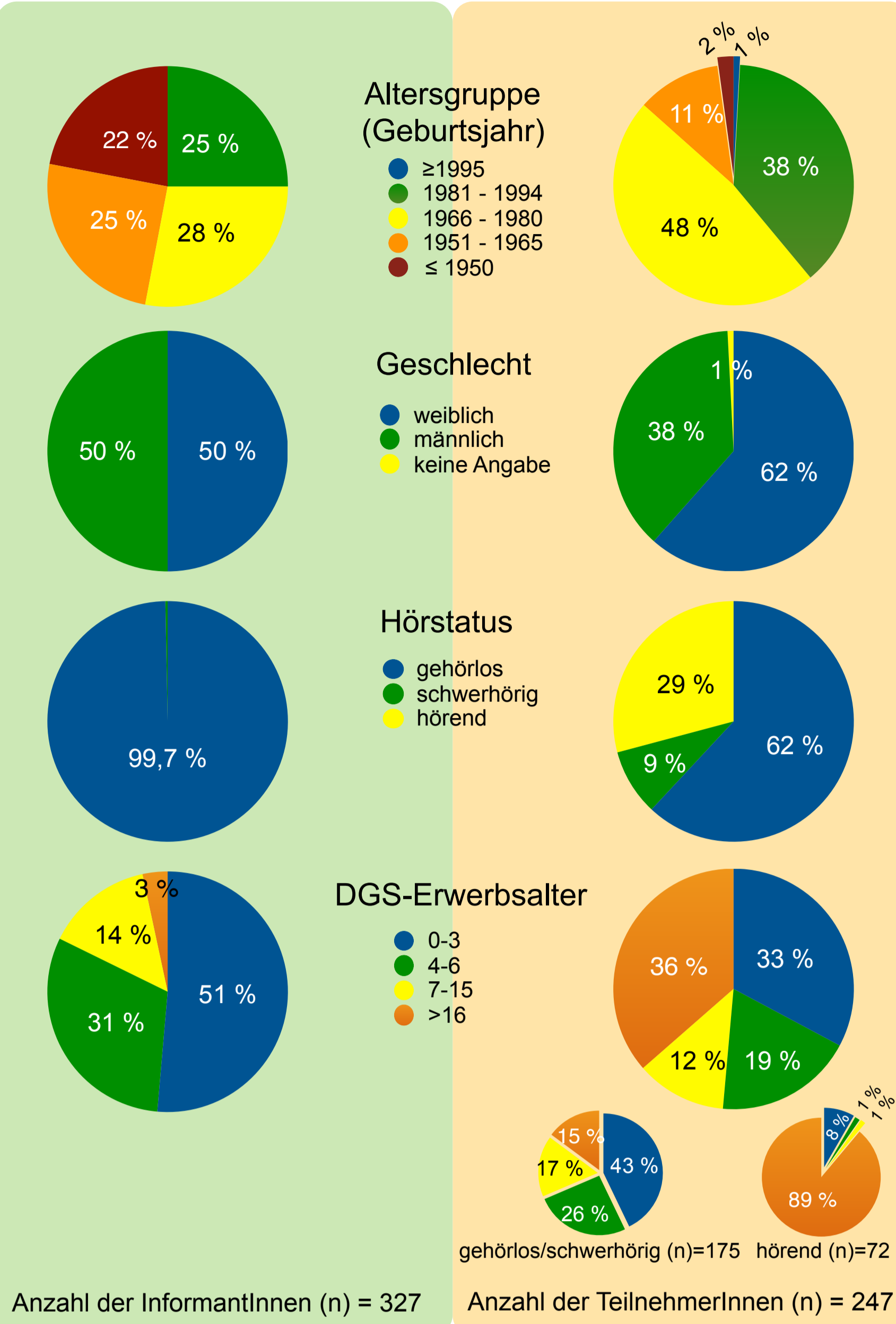
Variation lexikalischer Einheiten in der DGS

Wie Gebårdensprachlexikographie von der Kombination zweier Methoden profitieren kann: Die Ergänzung von Korpusdaten durch Crowd Sourcing

Gabriele Langer, Susanne König, Silke Matthes, Nele Groß, Thomas Hanke
Universität Hamburg, Institut für Deutsche Gebårdensprache und Kommunikation Gehörloser • www.dgs-korpus.de

DGS-Korpus

- InformantInnen**
- Kontrollierte Stichprobe: Ausgewogen bezüglich Alter, Geschlecht und Region
 - Muttersprachler oder frühes Erwerbsalter, verankert in der Gehörlosengemeinschaft
 - In der Region verwurzelt (mindestens 10 Jahre in derselben Region)
 - Aus rechtlichen Gründen keine minderjährigen Informanten (Geburtsjahr: ≥ 1995)
 - Anzahl der Informanten: 327 (gefílm: 330)
- Methode**
- Gefílmte Konversationen und „staged communicative events“
 - Multi-modales Korpus, lemmatisiert und erschlossen mithilfe von iLEx
- Daten**
- Natürlichsprachliche Gebårdten im Kontext
 - Die inhaltliche Abdeckung des Korpus ist weitgehend ungesteuert und zufällig
 - ≈ 616 Stunden relevantes Gebårdtenmaterial mit etwa 4 Mio. Tokens (geschätzt)
 - Bisher lemmatisiert (2015-12-09): ≈ 38 Stunden: 322,344 Tokens
- Lexikographischer Nutzen**
- Verwendung im Kontext: Lesarten, Sprachgebrauch, Kollokationen, Grammatik
 - Informationen über: Häufigkeit, typische Verwendungen, Distribution
 - Authentische Beispiele
- Besonderheiten und Anmerkungen**
- Daten von wenigen, aber sorgfältig ausgewählten Gebårdensprachnutzern
 - Niedrigfrequente Gebårdten und Gebårdtenverwendungen sind möglicherweise gar nicht oder in nicht ausreichender Anzahl für eine Analyse enthalten (vgl. Atkins/Rundell 2008, 57-61)
 - Fehlende Belege bedeuten nicht, dass es eine bestimmte Gebärde oder ein bestimmtes Phänomen nicht gibt
 - Zeitaufwändige Lemmatisierung
 - Lemmatisierung läuft noch: Regionen bisher nicht gleichmäßig abgedeckt
 - Zeitaufwändige Annotation und Analyse folgen auf die Lemmatisierung



DGS-Feedback

- TeilnehmerInnen**
- Unkontrollierte Stichprobe, freiwillige Teilnahme, abhängig vom Zufall
 - Gehörlose, Schwerhörige, Spätertaubte, CI-Nutzer, Hörende, große Bandbreite unterschiedlicher DGS-Sprachkompetenzen
 - Spricht eher computeraffine Personen an: wenig ältere TeilnehmerInnen
- Methode**
- Online Umfrage mittels DGS-Feedback System
 - Fragen sowohl in DGS als auch in Schriftdeutsch
 - Antworten zu isolierten Einzelgebårdten (Sprachbewusstsein)
 - Antworten: ‚benutze ich‘, ‚kenne ich‘ oder ‚kann ich nicht‘
 - Es können jederzeit Videokommentare abgegeben werden
 - Fragen zu mehreren verschiedenen Gebårdten werden zu Paketen zusammengestellt
 - Ergebnisse sind sofort verfügbar (deskriptive Statistiken, Karten zur Regionalität)
- Daten**
- Abdeckung ist steuerbar: Gebårdten und Lesarten werden nach Bedarf eingestellt
 - Bisher: > 2000 Gebårdten (≈ 1000 Formen mit ≈ 700 Lesarten) online, 152,528 Antworten von 247 TeilnehmerInnen (2015-12-14)
 - Positive („benutzt“) und negative Informationen („bekannt“ und „unbekannt“), sowie zum passiven Wortschatz („bekannt“), keine fehlenden Antworten zulässig
- Lexikographischer Nutzen**
- Bestätigung (oder Ablehnung) präsentierter Einheiten (z.B. Formen, Lesarten)
 - Größere Abdeckung der Verteilung (v.a. der regionalen), Ergänzung der Korpusdaten
 - Informationen zu Zitatformen, Ausführungsvarianten und weiteren Bedeutungen
- Besonderheiten und Anmerkungen**
- Grundgedanke: jede oder jeder aus der Sprachgemeinschaft kann sich beteiligen
 - In der Praxis: Anwerbung und längerfristige Bindung der TeilnehmerInnen ist nicht einfach
 - Umfrage läuft noch: laufende Anwerbung sowie Vorbereitung weiterer Gebårdten
 - Die Art der Stimulus-Präsentation kann Einfluss auf die Ergebnisse haben

Soziolinguistische Variation entdecken – Vorteile der Methodenkombination

Regionale Variation

Die sieben meistverwendeten Gebårdten für „Montag“ weisen regionale Verteilungsmuster auf.

144 Korpus-tokens von 54 InformantInnen und 108 ‚benutzt‘-Antworten von 78 Feedback-TeilnehmerInnen (gesamt: Daten von 131 Personen) wurden analysiert.

Einige Formvarianten wurden für die Auswertung zu einer Gebärde zusammengefasst. (2015-12-10)

Mögliche Angabe im Wörterbuch (beispielhaft): Die Gebärde MONTAG5 wird überwiegend in Süddeutschland verwendet.

Altersbedingte Variation

Im Korpus verwenden mehr Junge als Alte die Gebärde FILM. Dies mag daran liegen, dass jüngere Menschen eher an Themen interessiert sind, die mit Film und Filmen zu tun haben. Daten aus dem Feedback zeigen jedoch, dass der Anteil junger Menschen, die die Gebärde FILM in verschiedenen Lesarten verwenden, geringer ist als der älterer Menschen. Dies deutet möglicherweise auf beginnenden Sprachwandel in diesem Bereich hin. Frage: Welche Gebårdten verwenden Jüngere für diese Bedeutungen? Die Gebärde FILM bleibt weiter unter Beobachtung.

Mögliche Angabe in einem zukünftigen Wörterbuch: Lesarten „Schauspieler“, „Film (Fotoapparat)“, „Toilette“: veraltend.

Verwendung bei Frauen & Männern

FRAU deutet die Form der weiblichen Brust an. Manche Menschen finden diese Gebärde für „Frau“ unangemessen. Vermutung: Als Betroffene verwenden weniger Frauen diese Gebärde. Im Korpus findet sich nur ein geringer Unterschied: etwas weniger Frauen als Männer verwenden diese Gebärde, insgesamt ist sie aber weit verbreitet (441 Belege von 168 Personen). Im Feedback geben deutlich weniger Frauen als Männer an, FRAU (mind. 1 von 3 Lesarten: „weibliche Erwachsene“, „Lebenspartnerin“, „Anrede“) zu benutzen. Dabei nutzen Hörende (9 von 28: 32 %) die Gebärde deutlich seltener als Gehörlose/Schwerhörige (72 von 106: 68 %).

Mögliche Angabe im Wörterbuch: Die Gebärde ist weit verbreitet. Beachte: Die Gebärde wird z.T. als unangemessen empfunden, vor allem von Frauen.

Verdichtung regionaler Verteilung

Die Korpus-tokens der Gebärde SCHWARZ zeigen eine regionale Tendenz. Diese wird nicht nur durch die ‚benutzt‘-Antworten (gelb/orange) aus dem Feedback bestätigt und verdichtet, sondern auch durch die negativen Antworten von TeilnehmerInnen aus anderen Regionen (‚bekannt‘ und ‚unbekannt‘: grau). Regionen in gelb/orange können auch negative Antworten enthalten, aber da hier der Fokus auf der aktiven Verwendung liegt, werden diese nicht dargestellt.

Mögliche Angabe im Wörterbuch: regionale Gebärde (Süddeutschland), aber in ganz Deutschland bekannt.

Aktiver und passiver Wortschatz

Im DGS-Feedback wird die Verwendung von Gebårdten explizit erhoben. So erhalten wir auch Informationen über Gebårdten, die von Teilnehmenden zwar nicht benutzt werden, ihnen aber dennoch bekannt sind.

Mögliche Angabe im Wörterbuch: regionale Gebärde (Süddeutschland), aber in ganz Deutschland bekannt.

Ermittlung der Bedeutungen

Lesarten, die im Korpus nur wenig belegt sind, können über das Feedback getestet werden (hier „billig“). Lexikographischer Aspekt: Die Lesart „billig“ wurde nicht bestätigt und wird daher nicht im Wörterbucheintrag aufgenommen.

Die Lesart „Geld“ im Kontext mit typischen Nachbarn: GEBEN, BRAUCHEN, BEKOMMEN, VERDIENEN, VIEL, KEIN, BEZAHLEN, SPAREN

Lesarten von GELD	Korpus-tokens	Feedback ‚benutzt‘-Antworten (n=114)	Anteil Feedback ‚benutzt‘
Geld	342	107	94 %
finanziell	11	77	68 %
Preis (z.B. einer Ware)	19	63	55 %
teuer (hoher Preis)	51	20	18 %
billig (niedriger Preis)	1	1	1 %

Mögliche Angaben im Wörterbuch Lesarten: „Geld“, „finanziell“, „Preis (einer Ware)“, „teuer“
Kollokationen (für „Geld“): GEBEN, BRAUCHEN, BEKOMMEN, VERDIENEN, VIEL, KEIN, BEZAHLEN, SPAREN.

Für die lexikographische Beschreibung geschriebener Sprachen ist die Auswertung großer Korpora Standard. Auch in der Gebårdensprachlexikographie werden Korpora zunehmend genutzt. Das ist ein wichtiger und notwendiger Schritt. Da aber gebårdensprachliche Korpora erheblich kleiner sind als die geschriebener Sprachen, ist es hilfreich, das Bild durch andere Methoden zu ergänzen. Ergänzende Daten kann zum Beispiel die speziell hierfür entworfene Online-Umfrage liefern und dadurch tiefere Einblicke in die Verteilung von Gebårdten, zu ihren Varianten und ihren Bedeutungen ermöglichen. Dies gilt vor allem solange das DGS-Korpus noch nicht vollständig lemmatisiert ist. Aber nicht nur aufgrund der eingeschränkten Korpusgröße sondern auch aus anderen Gründen halten wir eine Kombination der Methoden für wertvoll, da jede ihre Vorzüge bei der Beantwortung bestimmter Fragestellungen hat und zum Gesamtbild des Gebårdtengebrauchs beitragen kann. Korpusdaten sind unverzichtbar für Informationen über tatsächliche Gebårdtenverwendung im Kontext, wie zum Beispiel Wortformen, Varianten, kontextuelle Bedeutungen, Kollokationen, grammatisches Verhalten und typische Konstruktionen. Sie können auch nach soziolinguistischen Aspekten wie Alter, Geschlecht, Region und Register oder anderen Einflüssen auf die Gebårdtenverwendung ausgewertet werden. Die Online-Umfrage kann solche Informationen ergänzen und verdichten, aber darüber hinaus bietet sie auch andere Informationen, wie zum Beispiel über den passiven Wortschatz. Wie die oben gezeigten Beispiele illustrieren, kann die lexikographische Arbeit von einer Kombination beider Methoden, Sprachdaten aus der Sprachgemeinschaft zu erheben, profitieren.

Literatur:

Atkins, B.T. Sue / Rundell, Michael (2008): *The Oxford Guide to Practical Lexicography*. Oxford, New York: Oxford University Press.

Hanke, Thomas / Storz, Jakob (2008): „iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography“. In: Crasborn, Onno / Ethimiou, Eleni / Hanke, Thomas / Thoutenhoofd, Ernst D. / Zwietersloot, Inge (eds.): *Construction and Exploitation of Sign Language Corpora*. 3rd Workshop on the Representation and Processing of Sign Languages. Paris: ELRA, pp. 64-87. (URL: www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf).

Kristoffersen, Jette H. / Troelsgaard, Thomas (2012): Integrating corpora and dictionaries: problems and perspectives, with particular respect to the treatment of sign language“. In: Workshop Proceedings, 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. Language Resources and Evaluation Conference (LREC) Istanbul, May 2012, ELRA, pp. 95-100. (URL: http://www.lrec-conf.org/proceedings/lrec2012/workshops/24_Proceedings_SignLanguage.pdf).

Langer, Gabriele (2012): „A Colorful First Glance at Data on Regional Variation Extracted from the DGS-Corpus: With a Focus on Procedures“. In: 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. Language Resources and Evaluation Conference (LREC) Istanbul, May 2012, pp.101-108. (URL: <http://www.lrec-conf.org/proceedings/lrec2012/index.html>).

Langer, Gabriele / König, Susanne / Matthes, Silke (2014): „Compiling a Basic Vocabulary for German Sign Language (DGS) – lexicographic issues with a focus on word senses“. In: Abel, Andrea / Vettorel, Chiara / Ralli, Natascha (eds.): *Proceedings of the XVI EURALEX International Congress: The User in Focus*, July 15-19 2014 in Bolzano/Bozen – Italy, pp.767-788. (URL: http://euralex2014.eurac.edu/en/callforpapers/Documents/EURALEX%202014_gesamt.pdf).

Matthes, Silke / Langer, Gabriele / Blank, Dolly / Hanke, Thomas / Konrad, Reiner / König, Susanne / Regen, Anja (2013): „Involving the crowd: How to complement corpus data in the process of dictionary making“. Abstract submitted for TISLR 11, July 10-13, 2013, London. (URL: http://www.ucl.ac.uk/dcall/abstracts/11r11_submission_239.pdf).

Nishio, Rie / Hong, Sung-Eun / König, Susanne / Konrad, Reiner / Langer, Gabriele / Hanke, Thomas / Rathmann, Christian. (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“. In: *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. (URL: <http://www.lrec-conf.org/proceedings/lrec2010/index.html>).

Diese Publikation wurde im Rahmen der gemeinsamen Forschungsförderung von Bund und Ländern im Akademiennprogramm mit Mitteln des Bundesministeriums für Bildung und Forschung und der Freien und Hansestadt Hamburg erarbeitet. Koordiniert wird das Akademiennprogramm von der Union der deutschen Akademien der Wissenschaften.