

Christine Möhrs/Carolin Müller-Spitzer (Mannheim)
**Bericht über die Methodenmesse im
Rahmen der IDS-Jahrestagung 2017**

1 Einleitung

Wie nun bereits seit einigen Jahren üblich, wurde die IDS-Jahrestagung auch dieses Jahr wieder von einer Methodenmesse begleitet, auf der sich passend zum Tagungsthema anwendungsorientierte Projekte mit Bezug zur Lexikonforschung präsentierten. Die Bandbreite der dargebotenen Themen war sehr groß: innovative methodische Ansätze im Bereich der Translationswissenschaft, Tools zur Analyse und Beschreibung lexikalischer Muster oder zur Detektion von Neologismen, neue lexikografische Ressourcen bis hin zu Infrastrukturaktivitäten und einem Kooperationsprojekt zwischen Schüler/innen und Wissenschaftler/innen zur Wortschatzanalyse. Im Folgenden sollen die einzelnen Projekte, die sich auf der Messe präsentiert haben, auf der Basis der eingereichten Abstracts der Messteilnehmer/innen kurz vorgestellt werden.

2 Präzentierte Projekte

Der Stand von Moritz J. Schaeffer und Silvia Hansen-Schirra (Johannes Gutenberg-Universität Mainz/Germersheim) beschäftigte sich mit dem Thema „**Übersetzungsentropie und Wortstellung: der Effekt auf Lesen und Schreiben**“. Bei diesem Forschungsthema wird in den Blick genommen, dass Elemente der Sprachen von Zweisprachlern koaktiviert werden (z.B. Macizo/Bajo 2006; Wu/Thierry 2012). Koaktivierung bedeutet, dass die Aktivierung eines semantischen oder strukturellen Elementes in einer Sprache auch gleichzeitig die Aktivierung des äquivalenten Elementes in der anderen Sprache hervorruft. Diese Koaktivierung kann von Vorteil sein, wenn man von einer Sprache in die andere übersetzt: Wenn schon beim ersten Lesen eines Wortes in der Ausgangssprache zielsprachliche Elemente koaktiviert werden, dann können diese auch schneller während der Produktion des Zieltextes verarbeitet werden. Allerdings kann die Koaktivierung auch hindernd sein, wenn die koaktivierten Elemente zu einem Text führen, der nicht akzeptabel ist im Sinne der Normen der Zielsprache. Dass Koaktivierung während der Übersetzung sowohl fördernd als auch hindernd sein kann, wurde behauptet,

allerdings lange nicht quantifiziert anhand des Verhältnisses zwischen den Sprachen während der Übersetzung. Carl/Schaeffer (2017) entwickeln zwei Metriken, die das zwischensprachliche Verhältnis während der Übersetzung quantifizieren. Am Stand im Rahmen der Jahrestagung wurden diese Forschungen erläutert und die „Translation Process Database“ (TPR-DB) präsentiert, die Prozessdaten (Augenbewegungen und Tastatureingaben) von mehr als 1.000 übersetzten Texten in mehr als 15 Sprachen beinhaltet. Die Augenbewegungen wurden mit einem Eyetracker aufgenommen. Sowohl Augenbewegungen als auch Tastatureingaben erlauben es, millisekundengenau zu beobachten, welchen Effekt z.B. Wortübersetzungsentropie auf das Lesen und Schreiben während der Übersetzung haben. Neben der Vermittlung der Thematik konnten die Besucher selbst an einem Eyetracker ihre Augenbewegungen beim Lesen und Schreiben aufzeichnen lassen.

Das einzelsprachenunabhängige Analyseprogramm „**lexpan – Lexical Pattern Analyser**“ wurde im IDS-Projekt „Usuelle Wortverbindungen“ zur Unterstützung der linguistischen Interpretation großer Sprachdatenmengen in Korpora entwickelt und am Stand von Annelen Brunner (IDS, Mannheim) präsentiert. Ursprünglich als heuristisches Werkzeug zur Ermittlung von Festigkeit und Varianz von usuellen Wortverbindungen (Steyer 2013) und für neuartige lexikografische Darstellungsformen (Steyer/Brunner 2009–2012) gedacht, ist es mittlerweile für die Rekonstruktion sprachlicher Verfestigung und Musterbildung generell einsetzbar (vgl. *lexpan*-Webseite). *lexpan* erlaubt es, große KWIC-Mengen gezielt zu strukturieren und qualitativ aufzubereiten. Ein zentrales Feature ist die Ermittlung von Füllern innerhalb von Muster-Slots auf der Basis gezielter Suchanfragen, die neben fest definierten Wortoberflächen Leerstellen enthalten, z.B. [*aus X (politischen/unbekannten/...) Gründen*], [*nach Belieben X und X (schalten und walten/an- und ausschalten/...)*], [*mit X X X (einem Hauch von/einer Prise von/...) Genugtuung*]. Die Füller können ausgezählt, in Frequenztabellen dargestellt und mit Kommentaren und Tags versehen werden, um sie z.B. nach semantischen und pragmatischen Gesichtspunkten zusammenzufassen. *lexpan* ermöglicht zudem die linguistische Kommentierung und Annotation von Kookkurrenzprofilen nach dem gleichen Prinzip sowie den Datenexport in XML- und HTML-Formate. Das Programm unterstützt damit sowohl die explorative, korpusgesteuerte Untersuchung von Mehrwortstrukturen und sprachlichen Mustern als auch neuartige Darstellungsformate.

Patrick Hanks (University of Wolverhampton) widmete sich an seinem Stand „**Three Types of Semantic Resonance**“ einer Studie, die sich mit der nicht-literalen Bedeutung von Wörtern und Phrasen beschäftigt. Ausgehend von der „Theory of Norms and Exploitations“ (Hanks 2013) haben Wörter keine Bedeutung als solche, sondern sie besitzen vielmehr Sinnpotenzial, das u.a. durch unterschiedliche Kontexte aktiviert werden kann. Die durch Kontext entstehenden Wortver-

bindungen sollen hier der Phraseologie zugerechnet werden, bei der aus einer Mischung aus Kollokationen und Valenzen die Bedeutung von Texten und Konversationen entsteht. Besonders in literarischen Texten liegen neben der normalen und idiomatischen Verwendung von Wörtern weitere Verwendungsweisen vor, die bei Wörtern nicht-wörtliche Bedeutungen erzeugen. Patrick Hanks zeigte an seinem Stand drei Formen, bei denen neue und originelle Bedeutungen durch irgendeine Art von Resonanz mit einem anderen, bereits etablierten sprachlichen Phänomen, geschaffen werden: 1) Lexikalische Resonanz (Metapher und andere Arten der figurativen Sprache), 2) Intertextuelle Resonanz und 3) Erfahrungsresonanz.

Wie die Bedeutung von Wörtern korpusbasiert untersucht und lexikografisch aufbereitet werden kann, haben Schüler/innen des Johann-Sebastian-Bach-Gymnasiums aus Mannheim in einem Kooperationsprojekt (Partner: IDS Mannheim, Lehrstuhl für Germanistische Linguistik der Universität Mannheim, das Johann-Sebastian-Bach-Gymnasium in Mannheim sowie die Albertus-Magnus-Schule in Viernheim) erlernt. Entstanden ist ein Wörterbuch: Das Denktionary ist ein wikibasiertes Wörterbuch, das die Schüler/innen aus drei neunten Klassen des Bachgymnasiums in Mannheim verfasst haben, z.B. zu Wörtern wie *vorglühen*, *Lückenfüller* oder *entgeistert*. Die Schüler/innen präsentierten auf der Messe das Wörterbuch und ihre Wörterbuchartikel am Stand **„Denktionary – Ein wikibasiertes Wörterbuch des Denkwerk-Projektes“** und wurden dabei von Antje Töpel und Nadja Radtke sowie Mentor/innen des Projektes begleitet. Im Projekt „Schüler machen Wörterbücher – Wörterbücher machen Schule“ (Förderung: Denkwerk-Programm der Robert Bosch Stiftung) lernten die Teilnehmenden einen empirisch-deskriptiven Blick auf die deutsche Sprache kennen, insbesondere in den Bereichen der Korpuslinguistik und der korpusbasierten Lexikografie. Sie beschäftigten sich mit der softwaregestützten Analyse von Textkorpora sowie mit offenen Wissensressourcen wie dem Wiktionary und der Wikipedia. Das erworbene Wissen wendeten die Schüler/innen dann gezielt an, um eigene Wörterbuchartikel zu ausgewählten Wörtern im wikibasierten Denktionary zu verfassen. Im Schuljahr 2017/18 wird das Projekt gemeinsam mit dem Viernheimer Gymnasium umgesetzt.

Die Aspekte Dynamik, Erweiterbar- und Veränderbarkeit vom Wortschatz standen beim Stand von Quirin Würschinger, Daphné Kerremans, Jelena Prokic, Hans-Jörg Schmid (Ludwig-Maximilians-Universität München) im Mittelpunkt: **„NeoCrawler. Erkennen und Beobachten lexikalischer Innovationen im Web“**. Ziel des Projekts ist es, die Diffusion von englischen Neologismen im Web zu beobachten, beteiligte Faktoren und deren Einflussgrößen zu ermitteln und zugrundeliegende Mechanismen lexikalischer Innovation zu modellieren. Dabei wird auf Längsschnittdaten aus früheren Phasen des Projekts (Kerremans 2015) sowie auf Fallstudien zum Einfluss von sozialen Medien und medienspezifischen

Verwendungsweisen (Würschinger et al. 2016) aufgebaut. Der NeoCrawler (Kerremans/Stegmayr/Schmid 2012), der am Messestand als Anwendung auch live ausprobiert werden konnte, dient dazu, Daten zum erstmaligen Auftreten und zur weiteren Verbreitung von englischen Neologismen im World Wide Web zu erheben. Zum einen durchsucht der Crawler das Web nach Kandidaten für lexikalische Innovation und speichert diese in einer Datenbank ab. Zum anderen sammelt er mithilfe der Google Custom Search API wöchentlich Daten zu deren Verbreitung und erstellt ein Korpus der Gebrauchskontexte. Dabei wird eine Reihe von Pre- und Post-Processing-Schritten zur Bereinigung, Vorbereitung und Analyse der Daten durchgeführt. Dadurch können Daten über die Gebrauchshäufigkeiten der (potenziellen) Neologismen, über spezifische Verwendungsweisen sowie über deren Verwendung in verschiedenen Texttypen und semantischen Domänen gewonnen werden, welche für die Diffusion lexikalischer Innovationen ausschlaggebend sind.

Am Stand **„Korpusbasierte Methoden zur Erstellung der LeGeDe-Resource“** wurden von Meike Meliss, Christine Möhrs und dem LeGeDe-Team die Arbeiten im Forschungsprojekt „Lexik des gesprochenen Deutsch (= LeGeDe)“ (IDS, Mannheim) dem Publikum präsentiert. Als Kooperationsprojekt der Abteilungen Lexik und Pragmatik am IDS hat das Projekt die Erstellung einer korpusbasierten elektronischen Ressource des gesprochenen Deutsch auf der Grundlage von Untersuchungen der Besonderheiten von mündlichem vs. schriftlichem Sprachgebrauch im Bereich des Lexikons zum Ziel (Leibniz-Wettbewerb 2016 der Leibniz Gemeinschaft, Förderlinie: Innovative Vorhaben). Die Untersuchungen zum Forschungsgegenstand dieses Projektes, der gesprochenen Sprache des Deutschen, erfolgen auf der Datengrundlage des „Forschungs- und Lehrkorpus gesprochenes Deutsch“ (FOLK; vgl. Schmidt 2014a, 2016; Kupietz/Schmidt 2015), das als erstes großes Gesprächskorpus des Deutschen mit einem besonderen Fokus auf gesprochener Sprache in Interaktion am IDS aufgebaut wird. Die „Datenbank gesprochenes Deutsch“ (DGD; vgl. Schmidt 2014b), über die FOLK recherchierbar ist, ist mit innovativen korpus technologischen Funktionalitäten ausgestattet und beinhaltet vielfältige Optionen der Erschließung mündlicher Daten nach linguistischen und interaktionalen Merkmalen. Im Rahmen der Methodenmesse wurden unterschiedliche empirische Methoden vorgestellt, die sich mit dem Gegenstandsbereich (typische lexikalische Einheiten der gesprochenen vs. geschriebenen Sprache) und den Eigenschaften der gesprochenen Lexik auf formaler, inhaltlicher und kommunikativ-funktionaler Ebene im Vergleich zu den Einheiten des geschriebenen Deutsch beschäftigen. Frequenzgesteuerte Daten zu Lemmata und Wortformen in Verbindung mit ihrer Kombinatorik (Bi- und Trigramme) und ihren Kookkurrenzprofilen sowie die detaillierte, semiautomatische Kodierung von zufallsgenerierten Stichproben zu ausgewählten Lemmata sind einige der

methodologisch relevanten Verfahren (Perkuhn/Keipel/Kupietz 2012; Seretan 2010). Sie werden sowohl für die Erstellung der Stichwortliste als auch für die Makro-, Mikro- und Mediostruktur der geplanten LeGeDe-Ressource, die langfristig in das „Online-Wortschatz-Informationssystem Deutsch“ (OWID) integriert werden soll, eingesetzt.

Thomas McFadden, Barbara Stiebels (Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin/Universität Leipzig) und Peter Meyer (IDS, Mannheim) präsentierten die **„ZAS-Datenbank zu satzeinbettenden Prädikaten“**, eine aktuell in OWID^{plus} (www.owid.de/plus/) veröffentlichte Ressource, in der auf der Messe am Stand direkt recherchiert werden konnte. Die ZAS-Datenbank dokumentiert das Satzeinbettungsverhalten von etwa 1700 Prädikaten des Gegenwartsdeutschen. In die Datenbank wurden für jedes Prädikat Korpusbelege aufgenommen, die das prädikatsspezifische Einbettungsverhalten im Hinblick auf die verschiedenen Satzeinbettungstypen belegen. Die Belege sind dann u.a. bezüglich Argumentstruktur und Argumentrealisierung des satzeinbettenden Prädikats sowie Eigenschaften der eingebetteten Struktur annotiert. Mit der Veröffentlichung der Datenbank in OWID^{plus} ist ein öffentlich zugängliches Werkzeug verfügbar, das den Inhalt der ZAS-Datenbank mit einer eigens eingerichteten Suchoberfläche verbindet, die es ermöglicht, komplexe Suchabfragen flexibel aufzubauen und schnell durchzuführen. Man kann sowohl nach Belegen für das Einbettungsverhalten spezifischer Prädikate, als auch nach den Prädikaten, die eine bestimmte Einbettungsstruktur lizenzieren, suchen, wobei auch kombinierte Abfragen (z.B. eingebettete Verbzweit-Sätze im Indikativ oder Prädikate, die Infinitivkomplemente, aber keine nominalisierten Komplemente zulassen) möglich sind.

Sascha Wolfer, Carolin Müller-Spitzer und Frank Michaelis (IDS, Mannheim) haben an dem Stand **„OWID^{plus} – eine neue Plattform für lexikalisch-lexikografische Ressourcen“** ergänzend dazu die gesamte Plattform OWID^{plus} vorgestellt, die im IDS-Projekt „Computerlexikografie“ seit 2015/2016 ergänzend zum Wörterbuchportal OWID entwickelt wird. OWID ist das Wörterbuchportal des IDS, in dem am IDS erarbeitete, korpusbasierte, wissenschaftliche lexikografische Ressourcen zum Deutschen eng vernetzt und in einer digital adäquaten Form aufbereitet zu finden sind. Mit OWID^{plus} wurde eine Plattform gegründet, in der auch experimentellere, datenzentrierte, multilinguale und in der Regel an Fachleute adressierte lexikalisch-lexikografische Ressourcen der Öffentlichkeit zur Verfügung gestellt werden können. OWID^{plus} ist wesentlich modularer organisiert als OWID, um mehr Flexibilität und Schnelligkeit bei der Integration neuer, auch IDS-externer Ressourcen zu ermöglichen. Über OWID^{plus} sind bereits fünf Rubriken verfügbar: *Wortschatzwandel in der ZEIT*, *Entropy Explorer* bzw. *Entropy data world map*, *Lexical change*, *Visits in the German Wiktionary* sowie die *ZAS-Datenbank*. OWID^{plus} möchte die Bandbreite der wortschatzbezogenen Angebote des

IDS erweitern und auch solchen Inhalten eine Plattform geben, die für die Forschungsgemeinschaft und universitäre Lehre sehr wertvoll sind, einem breiteren Publikum aber weniger zu vermitteln sind. Insofern ist das Ziel weniger, genau aufeinander abgestimmte Ressourcen zu veröffentlichen, sondern eher offen für vielseitige Angebote zu sein und daneben auch mit neuen Möglichkeiten der Visualisierung und Darstellung zu experimentieren. Außerdem dient OWID^{plus} im Sinne der reproduzierbaren Wissenschaft dazu, veröffentlichungs-begleitende Datensätze zur Verfügung zu stellen.

Schließlich stellten Thomas Gloning und Melanie Grunt Suárez (Justus-Liebig-Universität Gießen) mit einem Messestand **„Angebote für GermanistInnen in CLARIN-D“** vor. CLARIN-D ist das deutsche Teilprojekt des EU-Projekts CLARIN und steht für „Common Language Resources and Technology Infrastructure“. Das Ziel ist, Literatur-, Sprach- und Sozialwissenschaftler/innen mit digitalen Sprachdaten und digitalen Werkzeugen bei der Bearbeitung wissenschaftlicher Forschungsfragen zu unterstützen. Das Projekt wird vom Bundesministerium für Bildung und Forschung gefördert. In CLARIN-D wird eine Vielzahl von Sprachdaten für gesprochene und geschriebene Sprache angeboten, wie allgemeine Korpora aus dem IDS, aber auch Spezialkorpora z.B. zur Computer Mediated Communication (CMC) sowie historische Korpora. Ein weiteres Angebot ist die Dokumentation von Nutzungsszenarien. Die Facharbeitsgruppe „Deutsche Philologie“ hat begonnen, verschiedene wissenschaftliche Nutzungsszenarien in unterschiedlichen Darstellungsformen wie z.B. Anleitungstexte und Video-Experteninterviews zu dokumentieren. Auf der CLARIN-D Website und dem YouTube-Kanal können diese Darstellungsformen abgerufen werden. Drittens ist CLARIN-D daran interessiert, Sprachdaten von Literatur-, Sprach- und Sozialwissenschaftler/innen in die CLARIN-D Welt zu integrieren – seien das umfangreiche Projektdaten, einzelne Korpustexte, nützliche Werkzeuge oder digitale Anwendungsszenarien. Beispielsweise wird das Polytechnische Journal von Johann Gottfried Dingler als Teilkorpus im Deutschen Textarchiv zur Verfügung gestellt. Das letzte der vier Angebote informiert über digitale Werkzeuge und Webservices, mit denen Forscher/innen ihre Sprachdaten auswerten können. An insgesamt neun verschiedenen Universitäten und Institutionen sind CLARIN-D Zentren verortet. Das Institut für Deutsche Sprache in Mannheim ist eines dieser Zentren. Der Schwerpunkt liegt dort auf der deutschen Sprache, großen Korpora des Deutschen (nach 1900), Minderheitensprachen (Dialekte) sowie auf gesprochener Sprache. Vom IDS wird auch die Facharbeitsgruppe „Deutsche Philologie“ betreut.

3 Schlussbemerkung

Es hat sich bewährt, zusätzlich zu den meist eher theoretisch ausgerichteten Vorträgen auf der Jahrestagung eine Methodenmesse anzubieten, die zum einen sowohl IDS-externen als auch institutsinternen und eher anwendungsorientierten Projekten die Möglichkeit gibt, sich zu präsentieren. Außerdem bietet die Methodenmesse den Raum, sich intensiv mit Kolleg/innen auszutauschen. Wir freuen uns, dass die Messe auch dieses Jahr wieder viel Anklang gefunden hat und die Zeit über die intensiven fachlichen Gespräche wie im Flug vorbeiging. Ein Dank gilt insbesondere allen Teilnehmenden auch dafür, dass sie sich auf den „Elevator Pitch“, der Vorstellung der Projekte in 120 Sekunden im Plenum, so gut vorbereitet und sich alle an die restriktiven Zeitvorgaben gehalten haben.

Literatur

- Carl, Michael/Schaeffer, Moritz (2017): Measuring translation literality. In: Jakobsen, Arndt L./ Mesa-Lao, Bartolomé (Hg.): Translation in transition: Between cognition, computing and technology. Amsterdam, S. 81–105.
- Calzolari, Nicoletta et al. (Hg.) (2014): Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14). Reykjavik.
- Hanks, Patrick (2013): Lexical analysis: Norms and exploitations. Cambridge, MA.
- Kerremans, Daphné/Stegmayr, Susanne/Schmid, Hans-Jörg (2012): The NeoCrawler: Identifying and retrieving neologisms from the internet and monitoring ongoing change. In: Allan, Kathryn/Robinson, Justyna A. (Hg.): Current methods in historical semantics. (= Topics in English Linguistics (TiEL) 73). Berlin, S. 59–96.
- Kerremans, Daphné (2015): A web of new words. A corpus-based study of the conventionalization process of English neologisms. (= English Corpus Linguistics 15). Frankfurt a.M.
- Würschinger, Quirin et al. (2016): Using the web and social media as corpora for monitoring the spread of neologisms: The case of *rapefugee*, *rapeugee*, and *rapugee*. In: Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpiriST Shared Task. Berlin, S. 35–43.
- Macizo, Pedro/Bajo, M. Teresa (2006): Reading for repetition and reading for translation: Do they involve the same processes? In: Cognition 99, 1, S. 1–34.
- Kupietz, Marc/Schmidt, Thomas (2015): Schriftliche und mündliche Korpora am IDS als Grundlage für die empirische Forschung. In: Eichinger, Ludwig M. (Hg.): Sprachwissenschaft im Fokus. (= Jahrbuch des Instituts für Deutsche Sprache 2014). Berlin/Boston, S. 297–322.
- Perkuhn, Rainer/Keibel, Holger/Kupietz, Marc (2012): Korpuslinguistik. Paderborn.
- Schmidt, Thomas (2014a): The Research and Teaching Corpus of Spoken German – FOLK. In: Calzolari et al. (Hg.), S. 383–387.
- Schmidt, Thomas (2014b): The Database for Spoken German – DGD2. In: Calzolari et al. (Hg.), S. 1451–1457.

- Schmidt, Thomas (2016): Good practices in the compilation of FOLK, the Research and Teaching Corpus of Spoken German. In: Kirk, John M./Andersen, Gisle (Hg.): Compilation, transcription, markup and annotation of spoken corpora. Special Issue, International Journal of Corpus Linguistics 21, 3, S. 396–418.
- Seretan, Violeta (2010): Syntax-based collocation extraction. (= Text, Speech and Language Technology 44). Dordrecht u.a.
- Steyer, Kathrin (2013): Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht. (= Studien zur Deutschen Sprache 65). Tübingen.
- Steyer, Kathrin/Brunner, Annelen/Zimmermann, Christian (2013): Wortverbindungsfelder Version 3: Grund. <http://wvonline.ids-mannheim.de/wvfelder-v3/>.
- Wu, Yan J./Thierry, Guillaume (2012): Unconscious translation during incidental foreign language processing. In: NeuroImage 59, 4, S. 3468–3473.

Internetseiten bzw. -ressourcen

- CLARIN-D Facharbeitsgruppe „Deutsche Philologie“: www.clarin-d.de/de/facharbeitsgruppen/germanistik (Stand: 29.6.2017).
- CLARIN-D YouTube Kanal: www.youtube.com/user/CLARINGermany (Stand: 29.6.2017).
- CLARIN-D Zentren: www.clarin-d.de/de/aufbereiten/clarin-zentrum-finden (Stand: 29.6.2017).
- DGD = Datenbank gesprochenes Deutsch: <http://dgd.ids-mannheim.de/> (Stand: 29.6.2017).
- FOLK = Forschungs- und Lehrkorpus Gesprochenes Deutsch: <http://agd.ids-mannheim.de/folk.shtml> (Stand: 29.6.2017).
- lexpan* – Lexical Pattern Analyzer: www.ids-mannheim.de/lexik/uwv/lexpan.html (Stand: 29.6.2017).
- OWID^{plus}: www.owid.de/plus/ (Stand: 29.6.2017).