



Messe zur Technologie für mündliche Sprachkorpora

im Rahmen der IDS-Jahrestagung 2012

- Zeit:** Mittwoch, 14.3.2012, 10.00 Uhr bis 16.00 Uhr
- Ort:** im Kongresszentrum Rosengarten Mannheim, 1. Stock
(vor dem Musensaal)
- Organisation:** Thomas Schmidt (thomas.schmidt@ids-mannheim.de)
Wilfried Schütte (schuette@ids-mannheim.de)

Die Veranstaltung ist nach 2008, 2009 und 2011 mittlerweile die 4. Fachmesse dieser Art. Sie findet in diesem Jahr – wie schon 2009 – im Rahmen der IDS-Jahrestagung statt, die in diesem Jahr zu dem Thema „Das Deutsch der Migranten“ steht. Aussteller der Fachmesse sind Software-Entwickler und -Anbieter, Plattform- oder Datenbank-Betreiber, Korpus-Anbieter, Dienstleister und IT-Experten rund um die Erstellung und Auswertung von (mündlichen) Sprachkorpora. Passend zum variationslinguistischen Thema der IDS-Jahrestagung sind auch Ausstellerinnen und Aussteller vertreten, die zwar keine korpuslinguistischen Werkzeuge entwickelt, aber einschlägige Korpora erstellt haben.

After 2008, 2009 and 2011, this will be the 4th trade fair of its kind. As in 2009, it will be organized within the IDS Annual Conference ("German of migrants"). Exhibitors are software developers and vendors, platform or database operators, corpus providers, and IT experts engaged in the creation and evaluation of (oral) language corpora. Since the theme of this next annual meeting is variational linguistics, researchers who have created relevant corpora are present too.

Ausstellerinnen und Aussteller:

(Stand: 16.2.2012)

AGD – Archiv für Gesprochenes Deutsch	3
ANNIS – Across all layers: corpus querying and visualization using ANNIS	4
audiotranskription.de – Lösungen zur digitalen Aufnahme und Transkription	5
CLARIN-D – Web- und zentrenbasierte Forschungsinfrastruktur für Geistes- und Sozialwissenschaftler	5
Dh – Das Korpus <i>Deutsch heute</i>	7
DGD 2.0 – Datenbank für Gesprochenes Deutsch	8
ELAN – EUDICO Linguistic Annotator	9
EXMARaLDA und das Hamburger Zentrum für Sprachkorpora in CLARIN-D	10
Falko – Fehlerannotiertes Lernerkorpus des Deutschen als Fremdsprache	10
Feldpartitur – Multikodale Transkription von Videodaten	12
FOLK – Forschungs- und Lehrkorpus Gesprochenes Deutsch	13
FOLKER – Transkriptionseditor	14
GAIS – Gesprächsanalytisches Informationssystem	15
gi – Gesprächsanalyse interaktiv	16
KgSR – Sprachwandel im Ruhrgebiet – Dokumentation, Datensicherung, Analyse	17
KiDKo – Das KiezDeutsch-Korpus	19
LiS – Komponenten einer Forschungsinfrastruktur für linguistische Daten – die LiS-Projekte am Hamburger Zentrum für Sprachkorpora und am Institut für Deutsche Sprache	20
REDE – Deutscher Sprachatlas	21
TLA – The Language Archive	22



AGD – Archiv für Gesprochenes Deutsch

In der Linguistik gewinnen Sprachkorpora eine immer größere Bedeutung für die Forschung. Vor allem für Gesprächskorpora gilt jedoch, dass ihre Erstellung sehr aufwändig ist und umfassende technische und methodische Kenntnisse erfordert. Deshalb bietet das Archiv für Gesprochenes Deutsch (AGD) am IDS Mannheim vielfältige Unterstützung an, um die Erstellung und wissenschaftliche Auswertung von Gesprächskorpora in Forschung und Lehre zu fördern.

Zentrale Aufgabe des Archivs ist es, Korpora aus abgeschlossenen Forschungsprojekten zu übernehmen und für zukünftige Forschung und Lehre in der Wissenschaftsgemeinschaft zu erhalten. Auf diese Weise sind in 60 Jahren ca. 45 Korpora und Bestände an Tonaufnahmen mit über 5000 Stunden Gesamtdauer gesammelt worden.

Um den wissenschaftlichen Bedürfnissen noch umfassender Rechnung tragen zu können, werden zurzeit in der Abteilung Pragmatik zwei moderne Forschungskorpora aufgebaut. Zum einen ist dies das Korpus "Deutsch heute", das die Variation des standardnahen Deutsch, wie es heutzutage im deutschen Sprachraum gesprochen wird, auf der Basis von 170 Erhebungsorten dokumentiert. Zum anderen entsteht das "Forschungs- und Lehrkorpus Gesprochenes Deutsch" (FOLK), das einen breiten Querschnitt von Transkripten, Ton- und Videoaufnahmen aus unterschiedlichsten Gesprächstypen in deutscher Sprache umfasst. Beide Korpora werden über die "Datenbank Gesprochenes Deutsch" (DGD), die umfassend überarbeitet und auf eine neue ORACLE-Plattform umgesetzt wurde, online zur Verfügung gestellt.

Eine weitere wichtige Aufgabe des Archivs ist es, die wissenschaftliche Gemeinschaft bei der Erstellung von Gesprächskorpora zu beraten und Informationen aller Art rund um die Korpustechnologie im "Gesprächsanalytischen Informationssystem" (GAIS) anzubieten. Um Anbieter und Nutzer dieser Technologie miteinander ins Gespräch zu bringen, veranstaltet das AGD Fachmessen und Kolloquien. Darüber hinaus beteiligt es sich auch aktiv an der Entwicklungsarbeit, z.B. mit dem neuen Transkriptionseditor FOLKER, dem Metadaten-Schema für die aktuelle DGD (Version 2.0) und der Unterstützung von GAT 2.

Kontakt: Ulf-Michael Stift (pragmatikservice@ids-mannheim.de)

Webadresse: <http://agd.ids-mannheim.de>



ANNIS – Across all layers: corpus querying and visualization using ANNIS

ANNIS is a modern multi-layer corpus database, search tool and web based¹ visualization interface for corpora annotated on multiple linguistic layers. Being able to perform combined searches on different layers multiplies the usefulness of each annotation layer.

ANNIS allows users to query and display a diversity of annotations at once, such as morphological analysis, syntax trees (both phrase and dependency based), information structure, coreference, rhetorical structure etc. The system is independent of specific corpora, tagsets and linguistic theories. Additionally, ANNIS supports multimodal data, most recently extended to include streaming of aligned audio and video files while jumping to specific time frames by clicking on an aligned word, phrase or annotation.

Annotations are created using expert tools for the respective annotation task. For instance, EXMARaLDA or ELAN could be used for transcription and morphological analysis, Synpathy or TrEd for syntax, MMAX2 for coreference, and RSTTool for rhetorical structure; at the same time, annotations can also come from automatic tools, such as taggers or parsers. The data is then converted and merged using the converter framework Pepper (part of the SaltNPepper project²). Pepper is able to convert a range of source formats into the ANNIS format and to merge these different layers of annotation into one single corpus based on common textual layers.

Multi-layer corpora created in this way allow research on the interplay of independently annotated factors in discourse, formulating questions like *'Are constituents occurring in subject position taken up again more frequently than other arguments and when?' , 'Which syntactic cues are there for turn-giving?'*

At the software demo, we will provide the following multi-layer corpora for interactive exploration: KiDKo (GAT-transcriptions of dialogues annotated with POS, phrase structure, topological fields), FALKO (learner data, annotated with target hypotheses, POS, syntax),

¹ Both a server version and a stand-alone application (ANNIS Kickstarter) for local use and field work are available freely and open source under the Apache License 2.0 from <http://www.sfb632.uni-potsdam.de/~d1/annis>.

² SaltNPepper is open source under the Apache License 2.0 and freely available (see: <https://korpling.german.hu-berlin.de/saltnpepper/>)

TiGer (POS, syntax), TüBa-D/Z 6.0 (POS, syntax, topological fields, different coreference relations), OntoNotes 1.0 (POS, syntax, named entity type, coreference).

Kontakt: Amir Zeldes (Amir.Zeldes@rz.hu-berlin.de)
Florian Zipser (f.zipser@gmx.de)
Julia Ritz (jritz@uni-potsdam.de)
Thomas Krause (thomas.krause@alumni.hu-berlin.de)

Webadresse: <http://www.sfb632.uni-potsdam.de/~d1/annis>
<http://korpling.german.hu-berlin.de/saltpepper/>



audiotranskription.de – Lösungen zur digitalen Aufnahme und Transkription

audiotranskription.de zeigt Lösungen zur digitalen Aufnahme und Transkription von Interviews, Diktaten u.v.m. Auf unserer Website bieten wir die kostenfreien Transkriptionsprogramme f4 (Windows) und f5 (Mac), das kostenfreie Praxisbuch Transkription, umfassende Test- und Vergleichsberichte zu digitalen Rekordern und diverse Tutorials für den Arbeitsalltag an. Im dazu passenden Onlineshop lassen sich Interview-Sets und der mittlerweile recht bekannte, gelbe USB-Fußschalter auf Rechnung erwerben.

Kontakt: Dr. Thorsten Dresing, Thorsten Pehl (info@audiotranskription.de)

Webadresse: <http://www.audiotranskription.de/>



CLARIN-D – Web- und zentrenbasierte Forschungsinfrastruktur für Geistes- und Sozialwissenschaftler

Erhard Hinrichs und Kathrin Beck

Die enormen Fortschritte in der Informations- und Kommunikationstechnologie wirken sich zunehmend auf die Rahmenbedingungen und die Möglichkeiten des Forschens und Lehrens in fast allen wissenschaftlichen Disziplinen aus. Einzelforschern und Forschungsver-

bünden können große digitale Datenmengen mit entsprechenden Analysemethoden und –werkzeugen über Internetportale und digitale Archive zur Verfügung gestellt werden.

Der Wissenschaftsrat hat auf die Bedeutung von Forschungsinfrastrukturen besonders für die Geistes- und Sozialwissenschaften hingewiesen. Diese Empfehlung des Wissenschaftsrats schließt sich den Entscheidungen des European Strategy Forums on Research Infrastructures an, das in seine europäische Roadmap für Forschungsinfrastrukturen auch fünf geistes- und sozialwissenschaftliche Projekte aufgenommen hat. Auf der Grundlage dieser Empfehlungen und Entscheidungen fördert das Bundesministerium für Bildung und Forschung seit dem 1.Mai/1. Juni 2011 das Verbundprojekt „CLARIN-D: Web- und zentrenbasierte Forschungsinfrastruktur für Geistes- und Sozialwissenschaftler“ für einen Zeitraum von zunächst drei Jahren. Die CLARIN-D-Infrastruktur soll zur nachhaltigen Verbesserung der Bedingungen für die Forschung in den geistes-, sozial- und kulturwissenschaftlichen Disziplinen führen.

Auf der Basis virtueller Forschungsinfrastrukturen mit integrierten und interoperablen Ressourcen und Diensten lassen sich traditionelle Fragestellungen durch eine breitere Datengrundlage nicht nur empirisch besser absichern, sondern es können gänzlich neue Forschungsparadigmen entwickelt und neuartige wissenschaftliche Erkenntnisse erzielt werden. Der Nutzen für die adressierten Disziplinen konkretisiert sich durch die folgenden geplanten Maßnahmen und Dienste:

- Unterstützung von Forschern bei der Erstellung von neuen Ressourcen im Bezug auf Standards und Metadaten
- Trainingskurse und Ausbildungsmodule für text- und sprachbasierte Werkzeuge und Technologien
- Kuration und Verfügbarmachung relevanter Text- und Sprachressourcen über das Internet
- die Bereitstellung möglichst webbasierter Werkzeuge und Technologien, die von Forschern benötigt werden, um kollaborativ, disziplinübergreifend und institutionell unabhängig interagieren zu können
- größtmögliche Unterstützung von Forschern und ihren Forschungsprozessen aus allen geistes-, sozial- und kulturwissenschaftlichen Disziplinen bei zunehmend komplexer werdenden Arbeitsabläufen
- einen möglichst freien und ungehinderten Zugriff auf Ressourcen (Daten, Dienste) und auf eine technische Forschungsinfrastruktur (lokal, national, europäisch)

Das CLARIN-D-Konsortium besteht aus den Kernpartnern (Zentren) Berlin-Brandenburgische Akademie der Wissenschaften, Institut für Deutsche Sprache, Max-Planck-Institut für Psycholinguistik, Universität Hamburg, Universität Leipzig, Universität München, Universität des Saarlandes, Universität Stuttgart und Universität Tübingen (Projektkoordinator). Weiterhin wird eng mit sieben anerkannten Professoren verschiedener geisteswissenschaftlicher und kulturwissenschaftlicher Fachdisziplinen und mit drei Rechen-

zentren (Forschungszentrum Jülich, Rechenzentrum Garching, Gesellschaft für wissenschaftliche Datenverarbeitung (GWDG) Göttingen) kooperiert.

Ansprechpartner und wissenschaftlicher Koordinator des CLARIN-D Verbunds: Prof. Dr. Erhard W. Hinrichs, Seminar für Sprachwissenschaft, Eberhard Karls Universität Tübingen, Wilhelmstr. 19, 72074 Tübingen

Kontakt: Prof. Dr. Erhard W. Hinrichs (erhard.hinrichs@uni-tuebingen.de)
Kathrin Beck (kathrin.beck@uni-tuebingen.de)

Webadresse: <http://clarin-d.net/>

Deutsch heute

Dh – Das Korpus Deutsch heute

Der Stand präsentiert das IDS-Korpus *Deutsch heute*. Das Korpus umfasst im Zeitraum von 2006 bis 2009 systematisch erhobenes Datenmaterial zur gesprochenen deutschen Standardsprache aus den Ländern Europas, in denen Deutsch den Status einer Amts- und Unterrichtssprache hat. Es enthält Sprach- und Sozialdaten von 830 Sprechern, die im Hinblick auf die Analyse der intersituativen Variationspektren jeweils in wechselnden Sprechsituationen aufgenommen worden sind (vertikale Variationsdimension). Zudem eignet sich das Korpus durch eine formalisierte Aufnahmesituation und ein engmaschiges Ortsnetz insbesondere auch zur Beschreibung der sprachraumbedingten Variation im Sprechstandard (horizontale Variationsdimension). Der Stand bietet Informationen zum Korpusdesign, zur Datenaufbereitung und zum Annotations- und Analyseinstrument *praat* sowie zur Perspektive und zu Ergebnissen der Auswertung im Rahmen des IDS-Projekts *Variation des gesprochenen Deutsch*. Datenbeispiele werden in Form von Sprachkarten und im Rahmen des Ratespiels *Hör mal, wo der spricht* präsentiert, bei dem die Standbesucher ihre Kenntnisse der regionalen Variation des Deutschen erproben können.

Kontakt: Dr. Stefan Kleiner (kleiner@ids-mannheim.de)
Dr. Ralf Knöbl (knoebl@ids-mannheim.de)

Webadresse: http://www.ids-mannheim.de/prag/AusVar/Deutsch_heute/



DGD 2.0 – Datenbank für Gesprochenes Deutsch

Die Datenbank für Gesprochenes Deutsch (DGD 2.0) ist ein Korpusmanagementsystem im Archiv für Gesprochenes Deutsch (AGD). In die Datenbasis der DGD 2.0 werden am IDS erstellte bzw. hier archivierte Korpora übernommen. Die Datenbasis wird kontinuierlich erweitert.

Im Mittelpunkt der Systemarchitektur der DGD 2.0 steht eine objektrelationale XML-Datenbank (Oracle 11g), die eine schemabasierte, native Speicherung von Metadaten und Transkriptionen ermöglicht. Das Datenbanksystem unterstützt Volltextsuche und kontextsensitives Informationsretrieval sowie standardisierte, dynamisch generierte Präsentationen von XML-Dokumenten.

Externen Nutzern bietet die DGD 2.0 nach einmaliger Registrierung einen webbasierten Zugriff auf ausgewählte Teile der AGD-Bestände für die Verwendung in Forschung und Lehre. Die Auswahl der Daten für die externe Nutzung ist abhängig von der Zustimmung des jeweiligen Datengebers, der seinerseits über die entsprechenden Nutzungs- und Verwertungsrechte verfügen muss. Außerdem maßgeblich für die Auswahl sind bestimmte Schutzbedürfnisse des Archivs.

Die DGD 2.0 ermöglicht Volltextsuchen auf Metadaten und Transkriptionsdaten. Die Entwicklung strukturierter Recherchemöglichkeiten, deren Ergebnisse als Grundlage einer weiterführenden qualitativen oder quantitativen Analyse verwendet werden können, ist geplant.

Mit der Beta-Version stellen wir die DGD 2.0 zum ersten Mal der wissenschaftlichen Öffentlichkeit zu Testzwecken zur Verfügung. Mit dieser Veröffentlichung verfolgen wir das Ziel, die vorhandene Funktionalität der DGD 2.0 im Praxisbetrieb zu testen und erste Eindrücke vom potentiellen Benutzerkreis und den Benutzerbedürfnissen zu erhalten. Für Verbesserungsvorschläge und andere Rückmeldungen (an die Adresse dgd@ids-mannheim.de) sind wir daher sehr dankbar.

Bestandteil der Beta-Version ist auch der erste Teil von FOLK, dem "Forschungs- und Lehrkorpus gesprochenes Deutsch". Mit FOLK werden zum ersten Mal neuere texttechnologische Ansätze zur Erhebung und Bereitstellung mündlicher Korpora im Kontext von AGD und DGD erprobt. Die DGD 2.0 fungiert dabei als primäre Plattform sowohl für die Publi-

kation von künftigen Erweiterungen des FOLK-Korpus als auch für die (Weiter-)Entwicklung der Funktionalität für die korpuslinguistische Analyse solcher Gesprächsdaten.

Kontakt: Sylvia Dickgießer M.A., Joachim Gasch M.A., Dr. Thomas Schmidt
(dgd@ids-mannheim.de)

Webadresse: <http://dgd.ids-mannheim.de>



ELAN – EUDICO Linguistic Annotator

ELAN is a multimedia annotation tool that it is being developed and maintained by the “The Language Archive” department of the Max Planck Institute for Psycholinguistics in Nijmegen. It is available for Windows, Mac OS X and Linux, can be downloaded free of charge and is open source.

ELAN is a desktop tool for manually annotating audio and/or video recordings. It supports up to 4 videos per annotation document. It is a multi-layered annotation system: annotations are contained in tiers and the tiers can be part of a tier hierarchy (i.e. tiers can have depending tiers). The annotations contain Unicode text and the annotation documents are stored in XML files (EAF). It provides facilities for searching in a single file or in a user definable set of files (local corpus). Other multiple file operations are available as well allowing to create, convert, edit and export files en masse. Recently different „working modes“ have been added, each mode geared to a specific task in the transcription process. New are the Segmentation mode, for rapid creation of segments, and the Transcription mode, for entering text into existing annotations easily and conveniently.

ELAN is applied in a variety of research areas in linguistics (and beyond) such as sign language research, field linguistics, conversation analysis, gesture studies and multimodal interaction research.

At the Fachmesse ELAN is presented as a demo.

Kontakt: Han Sloetjes (han.sloetjes@mpi.nl)

Webadresse: <http://www.lat-mpi.eu/tools/elan/>

EXMARaLDA und das Hamburger Zentrum für Sprachkorpora in CLARIN-D

EXMARaLDA ist ein System zum Erstellen, Verwalten, Auswerten und Publizieren von Korpora gesprochener Sprache. Es wurde vom Juli 2000 an im Sonderforschungsbereich 538 Mehrsprachigkeit der Universität Hamburg entwickelt und wird seit 2011 unter anderem im Hamburger Zentrum für Sprachkorpora (HZSK) weitergepflegt. EXMARaLDA findet vornehmlich Anwendung in der Gesprächsforschung, in der Spracherwerbsforschung, in der Dialektologie und in der Phonetik/Phonologie. Zu den mit EXMARaLDA erstellten Korpora gehören beispielsweise das Korpus der gesprochenen Sprache im Ruhrgebiet, das METU Corpus of Spoken Turkish und das Korpus „Sprachvariation in Norddeutschland“.

Auch am SFB selbst wurde EXMARaLDA zum Erstellen und Aufbereiten mehrsprachiger Korpora verwendet, die nach dessen Ende am Hamburger Zentrum für Sprachkorpora vorgehalten und für die Verwendung in der CLARIN-D-Infrastruktur, in der das HZSK eines von neun Zentren in Deutschland ist, vorbereitet werden.

Das Angebot auf der Fachmesse umfasst zum einen Demonstrationen der Werkzeuge des EXMARaLDA Systems (Partitur-Editor, Corpus Manager, Analysewerkzeug EXAKT). Zum anderen werden Konzept und Datenbestände des Hamburger Zentrums für Sprachkorpora vorgestellt und schließlich erläutert, wie sich Ressourcen und Werkzeuge mittelfristig in der entstehenden CLARIN-D-Infrastruktur wiederfinden sollen.

Kontakt: Dr. Kai Wörner (kai.woerner@uni-hamburg.de)

Webadresse: <http://www.exmaralda.org/>
http://www.corpora.uni-hamburg.de/sfb538/de_overview.html



Falko – Fehlerannotiertes Lernerkorpus des Deutschen als Fremdsprache

Falko ist ein dynamisches Korpus geschriebener Zusammenfassungen und Aufsätze von Lernern des Deutschen als Fremdsprache. Daneben beinhaltet es ein Kontrollkorpus von

Texten deutscher Muttersprachler. Die Daten werden stark kontrolliert erhoben und mit Metadaten bezüglich der Erhebung, der Lerner, deren Sprachbiographie und der Weiterverarbeitung versehen.

Mit Lernerkorpora können Lernerfehler untersucht oder die Lerneräußerungen mit denen von Muttersprachlern kontrastiert werden. Im ersten Fall werden die ungrammatischen Äußerungen mit grammatischen Entsprechungen in Beziehung gesetzt. Im zweiten Fall werden alle Lerneräußerungen analysiert. Auch hierbei müssen die ungrammatischen auf grammatische Entsprechungen bezogen werden, damit sie mit einem einheitlichen Annotationschema beschrieben werden können. Diese Entsprechungen nennen wir Zielhypothesen (ZH), welche in Falko explizit annotiert werden. Dabei gibt es eine ZH, die nur eine minimale Korrektur von ungrammatischen Strukturen vornimmt. Zusätzlich gibt es linguistisch reichere ZHs. Das Stand-off Multi-Layer-Design, in dem Falko gespeichert ist, ermöglicht es dabei, konkurrierende Annotationsebenen darzustellen und das Korpus beliebig um weitere Ebenen zu erweitern. Die Suche nach Fehlern im Suchtool ANNIS2 (<http://www.sfb632.uni-potsdam.de/d1/annis/>) wird durch automatische Annotation von Abweichungen der ZH vom Lernertext ermöglicht.

Die Zielhypothesen können als Paralleltexte zum Lernertext behandelt und ebenso wie dieser mithilfe automatischer Tools weiterverarbeitet werden. So ist neben dem Lernertext auch jede Zielhypothese mit Wortart und Lemma annotiert.

In der aktuellsten Version liegt Falko auch als Baumbank vor. Hierfür wurde die minimale ZH mit Abhängigkeiten annotiert. Diese können ebenfalls in ANNIS2 problemlos durchsucht und dargestellt werden.

Falko ist damit das erste frei verfügbare Lernerkorpus mit tiefer syntaktischer Annotation und automatisch annotierten Satzfunktionen. Es stellt somit der Lernerkorpusforschung eine wertvolle Ressource zur Verfügung, die es ermöglicht, über die traditionelle Untersuchung oberflächennaher Merkmale hinaus Einblicke in die syntaktische Komplexität von Lerner-sprache zu gewinnen.

Kontakt: Prof. Dr. Anke Lüdeling (Anke.Luedeling@rz.hu-berlin.de)
Marc Reznicek (Marc.Reznicek@staff.hu-berlin.de)

Webadresse: <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko/standardseite/>

Feldpartitur – Multikodale Transkription von Videodaten

Christine Moritz

Während im Bereich der *Audioaufzeichnung* (z.B. von Interviews) Forschende auf verhältnismäßig etablierte Transkriptionspraktiken blicken können, zählt die Thematik der Transkription von *Videoaufzeichnungen* noch immer zu den Herausforderungen für die Sozialforschung. Denn Videodaten sind hochkomplexe, multikodale Daten³, die nicht allein das Sprechen der Akteure fixieren, sondern weitere Informationen zu Mimik, Gestik – aber auch Bild, Musik, Geräusch, Bewegung, Licht, Situation und Handlung etc.

Im BMWi-geförderten Projekt „Feldpartitur⁴“ wurde eine intelligente Software speziell für die Transkription von Videos entwickelt, um diese Lücke zu schließen. Neu am System Feldpartitur ist, dass Videodaten im Prozess der Transkription nicht mehr ausschließlich übersetzt werden in Textdaten (Kodewechsel), sondern auf zwei Achsen (X- und Y-Achse) *Prozessdaten* in ihrer *Gleichzeitigkeit* erfasst werden (Metapher der „Partitur“). Forschende wenden für diesen Vorgang unterschiedliche konventionelle Zeichen- und Symbolsysteme an, um – je nach Forschungsfrage – ihre Videodaten zu erfassen, so etwa das Zeichensystem der Musiknoten oder der Filmsprache. Die Feldpartitur reagiert auf diese Weise auf das vielbeschworene Problem der Erfassung der Komplexität der Videodaten und berücksichtigt deren Eigenschaften der Gleichzeitigkeit und Linearität einzelner Konstituenten. Die in dieser Weise als Diagramm sich aufspannende Datenmatrix dient auf der einen Seite als Visualisierung der im Video sich darbietenden Strukturen, auf der anderen als Grundlage

³ Transkripte, die nicht nur den Sprechtext, sondern verschiedene Dimensionen interaktiven Verhaltens erfassen, werden üblicherweise als "multimodal" bezeichnet. Der Begriff bezieht sich auf die Wahrnehmungsmodi („Sinnesmodalitäten“) der Interaktion der abgebildeten Akteure. Im Gegensatz zum Begriff der Multimodalität umfasst der Begriff der „Multikodalität“ den Modus der zu vermittelnden Informationen im Video, also das Bezogenheitsverhältnis zwischen Video und der/m jeweiligen RezipientIn. Denn Informationen im Video können mittels unterschiedlicher Codesysteme, etwa in Form von Texten, Bildern oder auch durch die Darbietung spezifischer Simulationen bereitgestellt werden.

⁴ Das Projekt wurde gefördert durch ein EXIST-Gründerstipendium des Bundesministerium für Wirtschaft und Technologie (BMWi) im Zeitraum 11/2010 bis 10/2011 aufgrund eines Beschlusses des Deutschen Bundestages

für weitere qualitative und quantitative Auswertung (z.B. Excel, SPSS, individuelle xml-Exporte oder pdf-Exporte für MaxQDA, atlas.ti, etc.)

Neben den inhaltlichen Neuerungen wurden technische Neuerungen im Projekt umgesetzt, um den hohen technischen Anforderungen gerecht zu werden. Die Software Feldpartitur wurde in der cloud-computing-Technology entwickelt: Für die Zeit während der Videotranskription wird für die jeweiligen Videoschnitte eine komfortable SSL-verschlüsselte SaaS-Software (Software-as-a-Service) genutzt. Dies ermöglicht neben der Entlastung der Forschenden von allen Konvertierungs- und Wartungsarbeiten internetbasiertes und somit system- und standortunabhängiges, überregionales sowie internationales Arbeiten in einem Team.

Literatur

Moritz, Christine: Videotranskription in der Qualitativen Sozialforschung: multidisziplinäre Annäherungen an einen komplexen Datentypus. Wiesbaden: VS-Verlag (i.V. 2012)

Moritz, Christine: [Die Feldpartitur. Multikodale Transkription von Videodaten](#). In: Uwe Flick, Ralf Bohnsack, Christian Lüders, Jo Reichertz. (Hrsg.): Reihe Qualitative Sozialforschung. Praktiken – Methodologien – Anwendungsfelder. Wiesbaden: VS-Verlag 2011.

Reichertz, Jo, Englert, Carina: [Einführung in die qualitative Videoanalyse. Eine hermeneutisch-wissenssoziologische Fallanalyse](#). In: Uwe Flick, Ralf Bohnsack, Christian Lüders, Jo Reichertz. (Hrsg.): Reihe Qualitative Sozialforschung. Praktiken – Methodologien– Anwendungsfelder. Wiesbaden: VS-Verlag 2010.

Kontakt: Dr. Christine Moritz (info@feldpartitur.de)

Webadresse: <http://www.feldpartitur.de>



FOLK – Forschungs- und Lehrkorpus Gesprochenes Deutsch

Während inzwischen verschiedene gut stratifizierte Korpora des geschriebenen Deutsch wissenschaftsöffentlich zugänglich sind, gibt es noch keine vergleichbare Sammlung des gesprochenen Deutsch. Mit FOLK baut das IDS ein kontinuierlich wachsendes Korpus auf, welches Gesprächsdaten aus unterschiedlichen Bereichen des gesellschaftlichen Lebens (Arbeit, Freizeit, Bildung, Medien) im deutschen Sprachraum via Internet zugänglich macht. Für viele linguistische und gesprächsanalytische Untersuchungen wird damit die Notwen-

digkeit entfallen, eigene Korpora zu erstellen. Kultur- und medienwissenschaftliche Untersuchungen können durch FOLK vielfältige Einblicke in die Realität der sozialen Kommunikation im Deutschland der Gegenwart gewinnen. Der germanistischen Hochschullehre im In- und Ausland und dem Unterricht im Fach DaF bietet FOLK Anschauungsbeispiele des heutigen gesprochenen Deutsch.

Alle Aufnahmen werden konform zu GAT 2 als Minimaltranskripte in literarischer Umschrift transkribiert. Die Transkripte werden dabei mit dem Ton aligniert, sodass zu jeder Transkriptstelle der zugehörige Audioausschnitt unmittelbar zur Verfügung steht. Jedes Gespräch verfügt über umfassende Metadaten, in denen die Gesprächsumstände und soziodemographischen Sprecherdaten dokumentiert werden.

FOLK wird der wissenschaftlichen Öffentlichkeit über die Datenbank Gesprochenes Deutsch (DGD) 2.0 zur Verfügung gestellt, die sich derzeit im öffentlichen Beta-Test befindet. Über die DGD können Dokumentationsdaten, Transkripte und Zusatzmaterialien zu FOLK eingesehen und über eine Volltextrecherche durchsucht werden. Erweiterungen von FOLK werden in den kommenden Jahren regelmäßig über die DGD veröffentlicht werden.

Kontakt: Jenny Winterscheid M.A. (winterscheid@ids-mannheim.de)

Webadresse: <http://agd.ids-mannheim.de/html/folk.shtml>



FOLKER **Transkriptionseditor**



FOLKER – Transkriptionseditor

Um Transkripte für das neue "Forschungs- und Lehrkorpus" (FOLK) schnell und komfortabel erstellen zu können und in der Datenbank DGD 2.0 optimal auswertbar und darstellbar zu machen, musste ein neuer Editor mit einem konsistenten Datenmodell und einem zugehörigen XML-basierten Datenformat entwickelt werden. Seit 2010 steht der Editor auf der Website des AGD (agd.ids-mannheim.de) zum kostenlosen Download zur Verfügung. Die aktuelle Version mit multilingualer Benutzeroberfläche ist 1.1. Entwickelt wurde das Programm von Thomas Schmidt (früher Hamburg, jetzt IDS) in enger Abstimmung mit der Abteilung Pragmatik am IDS.

In den Editor sind ein Audioplayer und eine Darstellung des Sprachsignals als Oszillogramm eingebaut. So lassen sich effizient Segmente zum Transkribieren auswählen und beim Transkribieren automatisch Zeitmarken für diese Segmente setzen. FOLKER überprüft segmentweise die Texteingabe auf Einhaltung der GAT-Konventionen und Integrität der

zeitlichen Strukturen (keine Überlappungen bei Segmenten eines Sprechers). Im Editor kann das Transkript wahlweise als Folge von Segmenten, als Partitur (in Anlehnung an die Visualisierung im EXMARaLDA-Editor) oder als Folge von Sprecherbeiträgen dargestellt werden, bei denen einem Sprecher zugeordnete und aufeinander folgende Segmente zusammengefasst werden. Diese Darstellungsmöglichkeiten passen zu einer sinnvollen Abfolge von Arbeitsschritten beim Transkribieren: Die Ersteingabe eines Rohtranskripts ohne Berücksichtigung der genauen Extension von Überlappungen findet zweckmäßiger in der Segment-Ansicht statt, in der Partitur-Ansicht können dann in einer Korrektur-Durchsicht die Feinheiten von Überlappungen korrigiert werden; die Beitrags-Ansicht dient dem abschließenden Korrekturhören bzw. -lesen.

Die Transkripte lassen sich in unterschiedlichen Formaten ausgeben. Mit dem zusätzlichen Werkzeug OrthoNormal können für Transkripte in literarischer Umschrift zusätzliche orthografische Versionen erstellt werden, mit denen die Suche in Gesprächsdatenbanken erleichtert wird. FOLKER ist interoperabel mit EXMARaLDA und ELAN.

An der Unterstützung des GAT-2-Basistranskripts (Syntaxkontrolle und Ausgabe) sowie an weiteren Verbesserungen für die Ersteingabe, Korrektur und Ausgabe von Transkripten wird derzeit gearbeitet.

Kontakt: Dr. Wilfried Schütte (schuette@ids-mannheim.de)

Webadresse: <http://agd.ids-mannheim.de/html/folker.shtml>



GAIS – Gesprächsanalytisches Informationssystem

Das Gesprächsanalytische Informationssystem GAIS ist ein wissenschaftliches Fachinformationssystem für die Arbeit in der Gesprächsforschung und für die wissenschaftliche Gemeinschaft. GAIS wurde in der Abteilung Pragmatik des Instituts für Deutsche Sprache (IDS) aufgebaut und wird dort inzwischen vom Archiv für Gesprochenes Deutsch betreut. Adressat von GAIS ist die wissenschaftliche Gemeinschaft von fortgeschrittenen Studierenden bis Lehrstuhlinhabern und über Fachgrenzen hinweg (Linguisten, Soziologen, Sprechwissenschaftler, Psychologen, Pädagogen). Nutzerbefragungen und Erfahrungen mit Service- und Beratungsanfragen an das AGD und an die Mailliste „Gesprächsforschung“ haben ergeben, welche Informationen in diesem Kreis immer wieder nachgefragt werden. Dazu gehören Informationen über die Gemeinschaft selbst (Neuigkeiten, Personen, Veranstaltungen, Stellenmarkt, Mailliste, Projekte, Korpora) als auch über ihre Arbeitsweise (Aufnahme- und

Korpustechnologie). GAIS bietet weiterführende Informationen in Form einer Bibliographie und einer Linksammlung an.

Texte zur Entwicklung von Standardlösungen und zur Vorstellung und Handhabung korpustechnologischer Instrumente für die Gesprächsforschung werden in GAIS publiziert und mit Schulungsangeboten unterstützt. Derzeit wird GAIS ausgebaut zu einem Handbuch zur Korpuserstellung in Form einer Online-Plattform.

GAIS umfasst u.a. drei Informationsbereiche:

- „Community“ mit Informationen aus der wissenschaftlichen Gemeinschaft (aktuelle Meldungen, Veranstaltungen, Sammlung von persönlichen Webseiten, Stellenmarkt und Informationen zu Maillisten, insbesondere zur Mailliste Gesprächsforschung mit gegenwärtig über 1200 Teilnehmern).
- Bibliographie zur Gesprächsforschung (BGF) mit inzwischen über 22.000 Einträgen und einer Sammlung mit relevanten Links.
- Technik mit Informationen zur Aufnahmetechnik (Audio, Video, Mikrofone, Kabel und Stecker), zur Korpusbearbeitung (juristische Fragen, Hard- und Software mit einem Tutorium zu Praat, Digitalisierung), Transkription (Konventionen und Editoren) und Korpusverwaltung (Datenbanken und Konvertierprogramme). Häufige Fragen werden im Bereich Korpusverwaltung unter FAQ beantwortet und technische Fachbegriffe unter „Fachbegriffe“ erklärt.

Kontakt: Dr. Wilfried Schütte (schuette@ids-mannheim.de)

Webadresse: <http://gais.ids-mannheim.de>



gi – Gesprächsanalyse interaktiv

„gi – Gesprächsanalyse interaktiv“ ist ein Blended-Learning-Angebot (Kombination von E-Learning und Präsenzlehre), das das Prinzip des kollaborativen forschenden Lernens einsetzt, um Studierende in die Gesprächsanalyse einzuführen. „gi“ wird seit September 2008 am Deutschen Seminar der Universität Zürich, Lehrstuhl Hausendorf, entwickelt und seit 2010 in der Lehre eingesetzt (Bachelor/Master).

Mit „gi“ wird eine Vermittlungsform etabliert, die den besonderen Anforderungen gerecht wird, die sich aus den methodologischen Grundannahmen der Gesprächsanalyse ergeben.

In „gi“ lernen die Studierenden die Bestandteile des gesprächsanalytischen Forschungsprozesses kennen, indem sie in begleiteten Arbeitsgruppen ein eigenes kleines Forschungsprojekt in all seinen Teilschritten durchführen. Mit seiner modularen Struktur bildet der Kurs den Forschungsprozess ab und setzt auf eigenes Forschen anstelle des reinen Nachvollziehens von Forschungsergebnissen. Ein wichtiger Mehrwert liegt in der größeren Nachhaltigkeit des Gelernten sowie einer in der Präsenzlehre kaum zu erreichenden Übersicht über den Forschungsprozess.

In „gi“ werden den Studierenden multimediale Lehrtexte mit authentischen Audio-/Videoaufnahmen zur Verfügung gestellt, mit denen sie theoretisches Wissen und praktische Fertigkeiten im Selbststudium erwerben können. Zentral ist jedoch etwas anderes: „gi“ nutzt die Möglichkeiten des Internets, um *kollaboratives* Forschen zu ermöglichen, und legt den Fokus auf interaktives Lernen. Für jede Forschungsphase stellt „gi“ daher Werkzeuge zur Verfügung, die das gemeinsame Arbeiten strukturieren (z.B. Wiki, Forum, Chat).

„gi“ verwendet die Lernplattform OLAT (*open source*) und kann daher auch außerhalb der Universität Zürich – sowohl im universitären als auch im außeruniversitären Bereich – eingesetzt werden. Die multimedialen Lerninhalte können auch in andere Lernplattformen integriert werden.

Kontakt: Katrin Lindemann M.A. (katrin.lindemann@ds.uzh.ch)
 Dr. Wolfgang Kesselheim (wolfgang.kesselheim@ds.uzh.ch)

Webadresse: <http://www.ds.uzh.ch/gi>



KgSR – Sprachwandel im Ruhrgebiet – Dokumentation, Datensicherung, Analyse

Projektgruppe KgSR – Ruhr-Universität Bochum

Die Projektgruppe KgSR (Korpus der gesprochenen Sprache im Ruhrgebiet) ist an der Ruhr-Universität Bochum in den Fachgebieten Germanistische Linguistik und Allgemeine Linguistik angesiedelt. Die Projektgruppe verfügt über ein größeres Korpus von Daten gesprochener (Ruhrgebiets-)Sprache aus den 80er-Jahren des 20. Jhs. (Alt-Korpus). Diese Daten

liegen in digitalisierter Form vor und werden z.Zt. aufwändig (unter Zuhilfenahme einer spezifisch an das Ruhrdeutsche angepasste EXMARaLDA-Transkription) transkribiert und annotiert; dazu wurden eigens spezifische Transkriptionskonventionen für das sog. ‚Ruhrdeutsche‘ entwickelt. Im Aufbau ist zudem die Erstellung eines aktuellen Vergleichskorpus (Neu-Korpus); die ersten Probeerhebungen werden voraussichtlich im Sommer d. J. durchgeführt.

Das Projekt 'Sprachwandel im Ruhrgebiet – Dokumentation, Datensicherung, Analyse' bündelt drei wesentliche Forschungsinteressen und -ziele. Neben der Dokumentation der ruhrdeutschen Varietät steht wesentlich die Analyse des Ruhrdeutschen im Fokus, und zwar unter der Fragestellung, inwiefern sich der in den letzten drei Jahrzehnten zu konstatierende soziale Wandel des Ruhrgebiets auch auf die gesprochene Sprache des Ruhrgebiets auswirkt, resp. ausgewirkt hat. Im Mittelpunkt der Analyse steht insofern die Frage, in welcher Form sich (neuere) Sprachwandelphänomene im Ruhrdeutschen finden und linguistisch erfassen lassen – dabei sind lexikologisch-lexikographische Fragestellungen ebenso relevant wie systemlinguistische (etwa phonologische und syntaktische). Zudem spielen Fragen der (sprachlichen) Auswirkungen von Migration ebenso eine Rolle wie demographische und/oder Gender-Perspektiven.

Ein drittes Projektziel betrifft die Nachhaltigkeit, resp. die Sicherung des umfangreichen Alt-Korpus und der neu zu erhebenden Daten. Gerade im Bereich der Geistes- und Kulturwissenschaften ist es bekanntlich besonders schwierig, Daten längerfristig zu konservieren – hier geht die KgSR-Projektgruppe in enger Zusammenarbeit mit der Universitätsbibliothek der Ruhr-Universität Bochum neue Wege der Korpusicherung und gleichzeitigen Veröffentlichung von Daten. Mittelfristig ist es ein Anliegen der Projektgruppe, sowohl die linguistische Analyse der Sprachwandelphänomene als auch die Erstellung des Neu-Korpus auf der Basis eines Drittmittelprojektes zu fördern.

Weitere Informationen zu diesem Projekt und zu den Projektmitarbeiter/inne/n sowie eine ausführliche Fachbibliographie, spezifische Transkriptionskonventionen sowie kleinere Tondokumente finden sich auf der Website des Projektes unter: rub.de/kgsr.

Kontakt: Dr. Kerstin Kucharzik (kgsr@rub.de)

Webadresse: <http://www.rub.de/kgsr>

kiezdeutsch

KiDKo – Das KiezDeutsch-Korpus

Heike Wiese, Ines Rehbein, Sören Schalowski, Ulrike Freywald, Katharina Mayr, Nadja Reinhold, Emiel Visser

(Universität Potsdam, Projekt B6 „Kiezdeutsch“ des SFB 632 „Informationsstruktur“)

Das KiezDeutsch-Korpus (*KiDKo*) ist entstanden im Rahmen des Teilprojekts B6 des Sonderforschungsbereichs 632 „Informationsstruktur“ der Universität Potsdam, der Humboldt-Universität zu Berlin und der Freien Universität Berlin. B6 befasst sich mit dem Sprachgebrauch Jugendlicher in multiethnischen urbanen Wohngebieten, dem sogenannten „Kiezdeutsch“ (vgl. Wiese et al. 2009). Das *KiDKo* umfasst spontansprachliche Daten aus Peer-Group-Gesprächen Jugendlicher im Alter von 14 bis 17 Jahren in Berlin. Die Daten liegen in Form von Audiodateien mit alignierten Transkriptionen vor; die Audiodateien wurden von den Jugendlichen bei Gesprächen in ihrer Freizeit in Abwesenheit von Erwachsenen oder gruppenfremden Personen (etwa Exploratoren) in Form von Eigenaufnahmen gewonnen. Die weitgehend auf Deutsch geführten Gespräche zwischen den Jugendlichen umfassen in erster Linie Face-to-Face-Kommunikationen, daneben auch einige Telefonate, und sind teils in öffentlichen, teils in privaten Räumen aufgenommen worden. Alle Anker-Sprecher/innen sind in Deutschland geboren und aufgewachsen.

Das Hauptkorpus umfasst 48 Aufnahmestunden (ca. 228.000 Token) und enthält Daten einer multiethnischen Population in Berlin-Kreuzberg; es basiert auf Aufnahmen von 17 Anker-Sprecher/innen unterschiedlicher (einschließlich deutscher) Herkunft aus einer Kreuzberger Schule, die 84,4 % Schüler/innen nicht-deutscher Herkunftssprache hat. Daneben umfasst das *KiDKo* ein Ergänzungskorpus mit 18 Aufnahmestunden (ca. 105.000 Token) von 6 Anker-Sprecher/innen deutscher Herkunft aus einer Hellersdorfer Schule mit 4,8 % Schüler/innen nicht-deutscher Herkunftssprache. Das Hellersdorfer Wohngebiet ist weitgehend monoethnisch, unterscheidet sich jedoch hinsichtlich weiterer sozioökonomischer Daten (Arbeitslosenquote, Anteil von Empfänger/innen von Existenzsicherungsleistungen u.ä.) nur minimal von dem multiethnischen Wohngebiet in Kreuzberg.

Die Aufnahmen sind im XML-Format transkribiert; die Transkriptionen folgen einer modifizierten Version der Konventionen des *Gesprächsanalytischen Transkriptionssystems* (GAT) für das Basistranskript (vgl. Selting et al. 1998). Für die Erstellung der Transkriptionen wurde das im SFB 538 „Mehrsprachigkeit“ der Universität Hamburg entwickelte Programm EXMARaLDA verwendet (*Extensible Markups Language for Discourse Annotation*; vgl. Schmidt & Wörner 2005). Um automatische Suchanfragen zu syntaktisch komplexeren

Phänomenen zu ermöglichen, ist eine weitere Aufbereitung der Daten vorgesehen, die mehrere Annotationsebenen liefern wird und es unter anderem ermöglicht, auf Wortarten, Phrasen und topologische Felder zuzugreifen. Das fertige Korpus soll mit Hilfe des Such- und Visualisierungswerkzeugs ANNIS (ANNotation of Information Structure) zugänglich gemacht werden. ANNIS wurde ebenfalls im Rahmen des Sonderforschungsbereichs 632 entwickelt (vgl. Zeldes et al. 2009) und ermöglicht komplexe Suchanfragen in Mehrebenen-Korpora mit vielfältigen Arten an linguistischer Annotation.

Durch seine umfangreiche spontansprachliche Datenbasis stellt das KiDKo eine neue empirische Ressource für Untersuchungen zum Sprachgebrauch in multi- und monolingualen urbanen Sprachgemeinschaften sowie generell zum informellen, dialogischen Sprachgebrauch unter Jugendlichen dar.

Kontakt: Prof. Dr. Heike Wiese (heike.wiese@uni-potsdam.de)

Webadresse: <http://www.kiezdeutsch.de/>



LiS – Komponenten einer Forschungsinfrastruktur für linguistische Daten – die LiS-Projekte am Hamburger Zentrum für Sprachkorpora und am Institut für Deutsche Sprache

Peter Fankhauser, Peter M. Fischer, Hanna Hedeland, Thomas Schmidt, Daniel Stein, Andreas Witt

Das DFG-Programm „**Literatur- und Informationsversorgungssysteme (LiS)**“ fördert Informationsdienstleistungen und innovative Informations-Infrastrukturen an den Hochschulen und Forschungseinrichtungen in Deutschland. Im Rahmen von LiS haben im vergangenen Jahr am **Hamburger Zentrum für Sprachkorpora (HZSK)** und im **Programmbereich Forschungsinfrastrukturen** des Instituts für Deutsche Sprache zwei Projekte ihre Arbeit aufgenommen, die sich den Aufbau wichtiger Komponenten einer Forschungsinfrastruktur für linguistische Daten zum Ziel gesetzt haben. Da die beiden Projekte aufgrund Ihrer ähnlichen Zielsetzungen – bei komplementärer Schwerpunktsetzung bezüglich der behandelten Datentypen – eng kooperieren, stellen wir sie bei der Fachmesse in einem gemeinsamen Angebot vor.

Das Projekt „Etablierung eines Schwerpunkts ‚Mehrsprachigkeit und gesprochene Sprache‘ im Hamburger Zentrum für Sprachkorpora“ zielt auf den Ausbau der computergestützten Methoden für die Arbeit mit Korpora gesprochener Sprache, die in einem Vorgängerprojekt (vornehmlich in Form des EXMARaLDA-Systems) entwickelt wurden. Zu diesem Zwecke wird erstens ein **Datenrepositorium** konzipiert und implementiert, über das existierende Datenbestände verwaltet, gepflegt und für eine Weiter- oder Wiederverwendung bereitgestellt werden. Zweitens werden verschiedene Instrumente für einen professionellen **Anwendersupport** entwickelt und erprobt.

Das Projekt „Zentrum für germanistische Forschungsprimärdaten“ am IDS Mannheim hat die Etablierung einer einheitlichen Infrastruktur für die nachhaltige Bereitstellung von Forschungsprimärdaten aus der germanistischen Linguistik zum Ziel. Hierfür soll ein **einheitliches Portal** aufgebaut werden, das die geschrieben- und gesprochensprachlichen Korpora sowie die elektronischen Lexika des IDS zusammenführt und einen zentralen Einstiegspunkt in diese Datensammlungen und ihre Metadaten bietet. Die Datengeber sollen dabei die Kontrolle über ihre Daten behalten und von weiteren **Langzeitarchivierungsfunktionen** des zugrunde liegenden Repositoriums (z.B. versionierbare und referenzierbare Momentaufnahmen) profitieren können.

Kontakt: Dr. Thomas Schmidt (thomas.schmidt@ids-mannheim.de)

Weblinks: LiS-Projekt am HZSK:
<http://www.corpora.uni-hamburg.de/lis/index.html>
Programmbereich Forschungsinfrastrukturen am IDS:
<http://www.ids-mannheim.de/fi/>



REDE – Deutscher Sprachatlas

Der Deutsche Sprachatlas hat mit dem Portal www.regionalsprache.de ein auf Sprachforschung ausgerichtetes Geoinformationssystem veröffentlicht. Dabei ist die Betrachtung von Sprachkarten im Rasterformat und aus einer Datenbank generierten Vektorkarten möglich. Der Benutzer erhält dadurch einen Einblick in die aktuellen Forschungsergebnisse des Deutschen Sprachatlas. Mit dem System können zudem eigene Karten erstellt und veröffentlicht werden.

Auf dem Portal sind unter anderem Ergebnisse der Projekte Digitaler Wenkeratlas, REDE, BayDat, PAD (Phonetischer Atlas Deutschlands) und SyHD (Syntax Hessischer Dialekte) dargestellt. Zudem werden zunehmend Rasterkarten in eine Vektordarstellung überführt.

Die Darstellung der Vektorkarten ist sehr flexibel. Unter anderem sind Vollformenkarten, symbolisierte Karten und die Anzeige von Statistiken im Raum mit Hilfe von Kuchendiagrammen möglich. Zusätzlich zu der Darstellung von Sprachkarten kann das System Tondokumente aus diversen Korpora, Wenkerbögen, sowie Fachliteratur als Karten darstellen und erlaubt unter anderem die geographische Suche in diesen Daten. Die Möglichkeit zur Einbindung neuer Datenquellen steht bei der Architektur des Systems im Vordergrund.

Das System nutzt zur Darstellung der Sprachkarten den Open-Source-Stack PostgreSQL/PostGIS, Umn-MapServer und OpenLayers. Als serverseitiges Framework wird ASP.Net verwendet. Die Datenbank PostgreSQL mit ihrer Erweiterung PostGIS bietet die Unterstützung für zahlreiche geometrische Operationen nahe an den Daten. Der Umn-MapServer dient als WMS für die Erstellung von Rasterkarten. Die Vektordaten werden durch Webhandler im kompakten GeoJSON- oder KML-Format vom Server an den Client übertragen. OpenLayers ist die clientseitige JavaScript Bibliothek, die für die Darstellung im Browser verantwortlich ist.

Kontakt: Slawomir Messner (slawomir.messner@staff.uni-marburg.de)

Webadresse: <http://www.regionalsprache.de/>



TLA – The Language Archive

The Language Archive (TLA) is a unit within the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands and is funded by the Max Planck Society, the Berlin Brandenburg Academy of Science and the Royal Dutch Academy of Science.

TLA has as one of its main goal to archive and preserve language resources from all over the world contributed by its funders' researchers as well as by external contributors. The archive currently contains more than 15.000 hours of recorded material from more than 200 different languages, many of which are threatened with extinction. The archive also contains several corpora of second language learners, some of which focus on migrant populations.

Besides being an archive, TLA also develops a number of linguistic tools such as the ELAN multimedia annotation tool and the LEXUS lexicon tool. TLA is very active in the area of digital research infrastructures for the humanities and plays a leading role in several projects funded by the European Commission to develop these.

At the IDS Fachmesse zur Korpustechnologie 2012 we will demonstrate the online access possibilities for The Language Archive.

Kontakt: Paul Trilsbeek (Paul.Trilsbeek@mpi.nl)

Webadresse: <http://www.mpi.nl/tla>