



Informationen zur Projektgruppe

Was bedeutet KgSR?

KgSR = Korpus der gesprochenen Sprache des Ruhrgebiets

An der Ruhr-Universität Bochum (Fachgebiete Germanistische Linguistik und Allgemeine Sprachwissenschaft) erarbeitet eine größere Gruppe von Linguistinnen und Linguisten zur Zeit sprachliche Spezifika des Ruhrdeutschen.

Dazu steht unser sog. ‚Alt-Korpus‘ der gesprochenen Sprache des Ruhrgebiets zur Verfügung. Mit der Datenerhebung für ein aktuelles Vergleichskorpus (‚Neu-Korpus‘) wird in diesem Jahr begonnen.

Die vollständige Erstellung des umfangreichen ‚Neu-Korpus‘ ist mittelfristig nur auf der Basis eines Drittmittelprojekts möglich. Das Projekt ist zur Zeit ohne Drittmittelförderung.

Projektmitglieder

Dr. Kerstin Kucharczik (Projektleitung), Katja Braunmüller, Jun.-Prof. Dr. Stefanie Dipper, Philipp Dorok M.A., Prof. em. Dr. Dr. h.c. mult. Siegfried Grosse, Prof. em. Dr. Dietrich Hartmann, Steffen Hessler B.A., Niels Kindl M.A., Jörg Kohlscheen M.A., Prof. em. Dr. Dr. h.c. Heinz H. Menge, André Posenau M.A., Daniel Pottmann B.A., Dipl.-Bibliothekar Ulrich Spiegelberg M.A., Jan Strunk M.A.

Datensicherung und öffentlicher Zugang

Doppelte Zielsetzung

1. Nachhaltige Sicherung der Daten des umfangreichen ‚Alt-Korpus‘ und des ‚Neu-Korpus‘
2. Gleichzeitig werden die Projektdaten für die wissenschaftliche Öffentlichkeit auf der Basis eines kennwortgeschützten Zugangs zur Verfügung gestellt.

Diese Ziele werden in Zusammenarbeit mit der Universitätsbibliothek der Ruhr-Universität Bochum verwirklicht.



‚Alt-Korpus‘

Dokumentation und Analyse

Das der Projektgruppe zur Verfügung stehende ‚Alt-Korpus‘ ist ein größeres Korpus von Daten der gesprochenen Sprache des Ruhrgebiets aus den 80er-Jahren des 20. Jahrhunderts.

Details zum ‚Alt-Korpus‘

- Es liegen ca. 120 Stunden Tonmaterial sowie Metadaten und Rohtranskripte von Interviews aus Schrebergärten vor.
- Das Tonmaterial ist komplett in digitalisierter Form vorhanden.
- Die Daten werden zur Zeit aufwändig transkribiert und annotiert.
- Es wurden eigene Transkriptionskonventionen für das ‚Ruhrdeutsche‘ erarbeitet.



‚Neu-Korpus‘

Erhebung und Zielsetzung

Die Erstellung eines aktuellen Vergleichskorpus (‚Neu-Korpus‘) befindet sich im Aufbau. Die ersten Probeerhebungen werden im Sommer 2012 durchgeführt.

Zielsetzungen bei der Datenerhebung

- Datenerhebung in Schrebergärten des Ruhrgebiets
- Thematische Schwerpunkte der Interviews orientieren sich am ‚Alt-Korpus‘
- Vergleichbare Probandengruppe
- Zusätzliche Probandengruppen: Frauen und Kinder

Zielsetzungen bei der Analyse

- Aufbau einer Longitudinalstudie
- Fragestellung: Inwieweit lassen sich (neuere) Sprachwandelphänomene im Ruhrdeutschen finden und systemlinguistisch und variationslinguistisch erfassen?

Unser linguistisches Interesse richtet sich insbesondere auf

- lexikologische Aspekte sowie lexikographische Auswirkungen
- phonologische Besonderheiten
- morphologische Aspekte
- syntaktische Untersuchungen
- soziolinguistische Fragestellungen (z.B. Genderfragen, Fragen der Migration)



Transkriptionskonventionen

Entwicklung eigener Transkriptionskonventionen

- Möglichkeit zur Darstellung aller im Ruhrdeutschen auftretenden phonologischen Phänomene
- Basierend auf HIAT (Rehbein & Schmidt 2004), GAT (Selting 2009) und DIDA (Schütte 2004)

Transkription in EXMARaLDA (Schmidt & Wörner 2009)

- EXMARaLDA bietet die Möglichkeit der Kombination von Primärdaten, Metadaten und Transkription.
- Die Verwendung standardisierter Datenformate (z.B. XML, UTF-8, Annotation-Graphs) gewährleistet die Nachhaltigkeit der Daten und ermöglicht die einfache Konvertierung in andere Formate, z.B. für die Präsentation als Webkorpus.
- Eigene virtuelle Tastatur in EXMARaLDA für das KgSR-Projekt



Segmentierung in Intonationseinheiten

- Erstellung einer segmentierten Transkription auf Basis der Prosodie und Pausenstruktur
- Das Zeichen ° markiert das Ende von Intonationseinheiten

Zwei Transkriptionsspuren

- Enge Transkription: Literarische Umschrift

- Verwendung der Standardorthographie
- Abweichungen von der Standardlautung oder typische „Merkmale des Ruhrdeutschen“ (Mihm 1997) werden mit standardorthographischen Mitteln und speziellen Sonderzeichen dargestellt.
- Annäherung an eine phonetische Transkription bei lesbarer Darstellung des Ruhrdeutschen
- Möglichkeit der gezielten Suche nach typischen Phänomenen der gesprochenen Sprache des Ruhrgebiets

- Weite Transkription: Normalisierte Umschrift

- Verwendung der Standardorthographie (inklusive Groß- und Kleinschreibung)
- Morphologische Normalisierung bei sprachlichen Ausdrücken, die im Standard als ungrammatisch gelten:

(1) vor vier Jahre
vor vier Jahren

- Basis für die (semi-)automatische Annotation (z.B. Wortarten-Tagging, Lemmatisierung, syntaktische Analyse)
- Die weite Transkription ermöglicht die einfachere Suche nach Varianten bestimmter Wörter oder Konstruktionen anhand der Standardorthographie.

Transkription phonologischer Besonderheiten

Markierung von abweichender Vokallänge

: für Länge, ː für Kürze (vgl. IPA)

(2) we:ch tʰɑch schøn
Weg Tag schon

Markierung von abweichender Vokalqualität

(3) speteːr ähˈlich
später ehrlich

Konsistente Markierung von Spirantisierung

(4) folchlich könich
folgich König

Sonorisierung

in Anlehnung an IPA markiert mit ͡

(5) iʃ=et gi͡b=et schlafanzuch
ist es gibt es Schlafanzug

Darstellung der <r>-Vokalisierung

- Die phonetische Transkription der Vokalqualität bei der <r>-Vokalisierung wäre sehr aufwändig und potentiell ungenau.
- Daher wird nur eine phonologische Markierung der <r>-Vokalisierung vorgenommen.

(6) duːch gaːˈten kindeːr
durch Garten Kinder

Markierung silbischer Konsonanten

mit einem vertikalen Strich ͡ (vgl. IPA)

(7) ihn͡n heizun͡g͡ schwim͡
ihnen Heizungen schwimmen

Kontraktion und Klitisierungen

Darstellung mit dem in Glossen üblichen Zeichen =

(8) dɔ=nɔ=mə da=s au=m schɔ=mə

