

Die Sprache

für Theorie, Praxis,
ation

011)

weise:

NSSTITUTS FÜR DEUTSCHE SPRACHE
/ON

3 M. EICHINGER
ng)
ische Sprache

sim

KA LINKE
ch
ar
9

irat:

F DEPPERMAN (Mannheim)
r HOFFMANN (Dortmund)
t (Mannheim)
KA REDDER (Hamburg)
IE VACHKOVÁ, Ph.D. (Prag)
ds des Internationalen Wissenschaft-
GS
eim.de/org/gremien/intwrat.html)

tsche Sprache

sim
annheim.de

verfahren:
elte Begutachtung (Peer Review)

erlag GmbH & Co. KG
0 G, D-10785 Berlin-Tiergarten
0 85-620, Fax: +49 30 25 00 85-305
rmedien.de

erlag GmbH & Co. KG
0 G, D-10785 Berlin-Tiergarten
25 00 85-223, Fax: +49 30 25 00 85-275
@ESVmedien.de

G, Kto.-Nr. 512 203 101, BLZ 100 708 48,
0848 0512 2031 01, BIC(SWIFT): DEUTDEB110

ingen:

1 im Abonnement: Jahresabonnementspreis
.. eJournal und Archiv); Einzelbezug je Heft
... für Studenten gegen Vorlage einer

preis € (D) 77,40; jeweils einschließlich 7 % Mehrwertsteuer
und zuzüglich Versandkosten. Die Bezugsgebühr wird
jährlich im Voraus erhoben. Abbestellungen sind mit einer
Prist von 2 Monaten zum 1.1. j.J. möglich. Keine Ersatz- oder
Rückzahlungsansprüche bei Störung oder Ausbleiben durch
höhere Gewalt oder Streik. Preise für gebundene Ausgaben
früherer Jahrgänge auf Anfrage.

Anzeigen:

Erich Schmidt Verlag GmbH & Co. KG
Genthiner Str. 30 G, D-10785 Berlin-Tiergarten
Tel: +49 30 25 00 85-621, Fax: +49 30 25 00 85 -305
e-mail: S.Valipour@ESVmedien.de
Anzeigenleitung: Sabine Valipour
Es gilt Anzeigenpreisliste Nr. 26 vom 1.1.2012, die auf
Wunsch zugesandt wird.

Manuskripte:

Von Text und Tabellen erbitten wir neben einem sauberen
Ausdruck auf Papier – möglichst ohne handschriftliche
Zusätze – das Manuskript in elektronischer Form bevor-
zugt in Word, sonst zusätzlich im RTF-Format. Zur Ver-
öffentlichung angebotene Beiträge müssen frei sein von
Rechten Dritter. Sollten sie auch an anderer Stelle zur Ver-
öffentlichung oder gewerblichen Nutzung angeboten wor-
den sein, muss dies angegeben werden. Mit der Annahme
zur Veröffentlichung überträgt der Autor dem Verlag das
ausschließliche Verlagsrecht und das Recht zur Herstellung
von Sonderdrucken für die Zeit bis zum Ablauf des Urheber-
rechts. Das Verlagsrecht umfasst auch die Rechte, den Bei-
trag in fremde Sprachen zu übersetzen, Übersetzungen zu
vervielfältigen und zu verbreiten sowie die Befugnis, den
Beitrag bzw. Übersetzungen davon in Datenbanken ein-
zuspeichern und auf elektronischem Wege zu verbreiten
(online und/oder offline), das Recht zur weiteren Vervi-
elfältigung und Verbreitung zu gewerblichen Zwecken im
Wege eines fotomechanischen oder eines anderen Verfah-
rens sowie das Recht zur Lizenzvergabe. Dem Autor verbleibt
das Recht, nach Ablauf eines Jahres eine einfache Abdruck-
genehmigung zu erteilen; sich ggf. hieraus ergebende
Honorare stehen dem Autor zu. Bei Leserbriefen sowie bei
angeforderten oder auch bei unaufgefordert eingereichten
Manuskripten behält sich die Redaktion das Recht der
Kürzung und Modifikation der Manuskripte ohne Rück-
sprache mit dem Autor vor.

Rechtliche Hinweise:

Die Zeitschrift sowie alle in ihr enthaltenen einzelnen
Beiträge und Abbildungen sind urheberrechtlich geschützt.
Jede Verwertung, die nicht ausdrücklich vom Urheberrechts-
gesetz zugelassen ist, bedarf der vorherigen Zustimmung
des Verlages. Das gilt insbesondere für Vervielfältigungen,
Bearbeitungen, Übersetzungen, Mikroverfilmungen und
die Einspeicherung und Verarbeitung in elektronische
Systeme. – Die Veröffentlichungen in dieser Zeitschrift ge-
ben ausschließlich die Meinung der Verfasser, Referenten,
Rezensenten usw. wieder. – Die Wiedergabe von Gebrauchs-
namen, Handelsnamen, Warenbezeichnungen usw. in dieser
Zeitschrift berechtigt auch ohne Kennzeichnung nicht zu
der Annahme, dass solche Namen im Sinne der Markenzei-
chen- und Markenschutzgesetzgebung als frei zu betrachten
wären und daher von jedermann benutzt werden dürfen.

ISSN: 0340-9341

Druck:

Danuvia Druckhaus Neuburg, Neuburg/Donau

Heike Wiesel/Ulrike Freywald/Sören Schalowski/Katharina Mayr

Das KiezDeutsch-Korpus

Spontansprachliche Daten Jugendlicher aus urbanen Wohngebieten

Abstract

Der Beitrag beschreibt ein Korpus spontansprachlicher Daten von Jugendlichen aus urbanen Wohngebieten und präsentiert Ergebnisse darauf basierender Korpusanalysen. Die Daten stammen aus Selbstaufnahmen von Gesprächigen Berliner Jugendlicher (23 Ankersprecher/innen, 14 bis 17 Jahre alt). Das Korpus umfasst rund 333.000 Token. Den größeren Teil stellen Daten ein- und mehrsprachiger Sprecher/innen dar, die aus einem Wohngebiet mit einem hohen Bevölkerungsanteil mit Migrationshintergrund stammen. Wie an anderer Stelle argumentiert, kann die multiethnische Jugendsprache, die hier gesprochen wird, das sog. „Kiezdeutsch“, als neuer, dynamischer Dialekt des Deutschen charakterisiert werden. Ein kleinerer Teil des Korpus bringt Daten aus einem sozioökonomisch vergleichbaren, aber weitgehend monoethnischen Wohngebiet zusammen. Das Korpus stellt eine neue empirische Ressource für Untersuchungen zum Sprachgebrauch in mehrsprachigen Kontexten, urbaner Jugendsprache und spontaner, informeller dialogischer Sprache dar. Wir illustrieren dieses Potenzial mit Korpusanalysen zu zwei Phänomenen auf Wort- bzw. Satzebene: die Verwendung von *so* als Fokusmarker und nicht-kanonische „Adv SVO“-Linearisierungen. Daneben geben wir eine Übersicht über das Korpusdesign und diskutieren Probleme der Transkription natürlicher Sprachdaten.

This paper describes a corpus of spontaneous speech by young people in an urban area of Germany and presents linguistic analyses drawing on this resource. The data are based on self-recordings of peer group conversations among adolescents in Berlin (23 anchor speakers aged 14 to 17). The corpus consists of approx. 333,000 tokens. Its main part draws on data from monolingual and multilingual speakers of German, all living in a neighbourhood with a high proportion of people with a migrant background. As is argued elsewhere, the multiethnic youth language spoken here, “Kiezdeutsch”, can be regarded as a new, dynamic German dialect. A smaller part of the corpus contains data from a largely monoethnic neighbourhood with comparable socioeconomic indicators. The corpus provides a new empirical source for studies of language use in multilingual contexts, urban youth language, and spontaneous, informal dialogic speech in general. We illustrate this potential with corpus analyses of two phenomena at the word and sentence levels: the use of *so* ‘such’ as a focus marker and non-canonical Adv SVO serialisations. In addition we give an outline of the technical corpus design and touch on problems of adequate methods of transcription for natural language data.

1. Einleitung

Im Vergleich zu schriftsprachlichen elektronischen Korpora sind gesprochensprachliche Korpora des Deutschen bislang immer noch selten,¹ und dies gilt vor allem für den Bereich der Spontansprache, der für die linguistische Analyse realen Sprachgebrauchs besonders interessant ist, jedoch eine große Herausforderung für die Datengewinnung sowie Korpuserstellung und -aufbereitung darstellt. Das im Folgenden vorgestellte KiezDeutsch-

¹ Für eine Übersicht über Korpora gesprochener Sprache im Netz (u.a. auch für das Deutsche) vgl. Merkel/Schmidt (2009). Beispiele für gesprochensprachliche Korpora des Deutschen sind die Korpora der „Datenbank Gesprochenes Deutsch (DGB)“ und das „Forschungs- und Lehrkorpus gesprochenes Deutsch (FOLK)“ am IDS Mannheim (Übersicht über die Korpora des Archivs für Gesprochenes Deutsch am IDS: <http://agd.ids-mannheim.de/html/korpora/korpus-index.shtml>), die „Tübinger Baumbank des

Korpus (KiDKo) bedient diesen Bereich und liefert hierbei jugendsprachliche Daten aus Peer-Group-Gesprächen. Speziell Korpora gesprochener deutscher Jugendsprache existieren bislang nur vereinzelt² und sind oft nur teilweise transkribiert und/oder nicht öffentlich zugänglich.

Das KiezDeutsch-Korpus konzentriert sich auf eine jugendsprachliche Varietät des Deutschen, die sich in multiethnischen urbanen Wohngebieten herausgebildet hat.³ Der Begriff „Kiezdeutsch“, der für diese Varietät verwendet wird, vermeidet negative Stereotype, wie sie von Bezeichnungen wie „Kanak Sprak“⁴ u.ä. evoziert werden, und betont die Zugehörigkeit der Varietät zum Bereich des Deutschen (im Gegensatz zur Charakterisierung als eigene Sprache/„Sprak“), schränkt sie dabei jedoch nicht auf Sprecher/innen eines bestimmten sprachlichen oder ethnischen Hintergrunds ein (anders als z.B. „Türkendeutsch“), sondern erfasst eine sprachlich und ethnisch vielfältige Sprechergemeinschaft mit ihrem informellen Sprachgebrauch im alltäglichen urbanen Wohnumfeld (Berlinisch „Kiez“).

Urbane Populationen mit ihrer besonderen Dynamik, ihrer sozialen, ethnischen und sprachlichen Mischung und den sie prägenden vielfältigen Prozessen der Zu- und Abwanderung sind für die Untersuchung sprachlicher Variation und sprachlichen Wandels besonders interessant, aus soziolinguistischer Perspektive ebenso wie aus grammatiktheoretischer Sicht. Dies gilt bereits für herkömmliche Dialekte; in noch stärkerem Maße schließlich für die neuen Stile und Dialekte, die sich, verbunden mit Migration und Mehrsprachigkeit, in den letzten Jahrzehnten unter Jugendlichen in urbanen Wohngebieten Europas ausgebildet haben und ein besonderes sprachliches Innovationspotenzial in sich bergen. Die sie tragenden multiethnischen und multilingualen Sprachgemeinschaften stellen allein quantitativ nicht länger einen marginalen Sonderfall dar, sondern sind gesellschaftlich ebenso wie sprachlich von steigender Bedeutung. Während in Deutschland heute bereits knapp ein Fünftel der Bevölkerung einen Migrationshintergrund aufweist, wird dies in Zukunft wegen des vergleichsweise hohen Anteils an Jugendlichen und Kindern in dieser Bevölkerungsgruppe noch zunehmen. So ist jede/r zweite Einwohner/in mit Migrationshintergrund jünger als 34,2 Jahre, gegenüber 46,5 Jahren bei der Bevölkerung ohne Migrationshintergrund. In der Altersgruppe der unter 5-jährigen Kinder weist bereits ein Drittel einen Migrationshintergrund auf, wobei in urbanen Gebieten der Anteil noch erheblich höher liegt und oft bereits über die Hälfte der Kinder ausmacht.⁵ Wir haben es hier also nicht mit einem Randphänomen zu tun, sondern mit einem zentralen Bereich, der im urbanen Raum in absehbarer Zeit den Normalfall bilden wird.

² So z.B. das Korpus „Jugendspezifische Sprechweisen“ (Schlobinski/Kohl/Ludewigt 1998) oder das an der Freien Universität Berlin erstellte JUSPiL-Korpus (Jugendsprache im Längsschnitt), vgl. http://www.geisteswissenschaften.fu-berlin.de/v/jugendsprache/Das_Korpus/Das_JuSpiL-Korpus.html.

³ Vgl. Wiese/Freywald/Mayr (2009). Das KiezDeutsch-Korpus ist entstanden im Rahmen des Teilprojekts B6 „Kiezdeutsch“ (Projektbeginn: 2008, Projektleitung: H. Wiese) des DFG-geförderten Sonderforschungsbereichs 632 „Informationsstruktur“ der Universität Potsdam, der Humboldt-Universität zu Berlin und der Freien Universität Berlin.

⁴ Der Versuch Feridun Zaimoglus, der mit seinem 1995 erschienenen Roman *Kanak Sprak* den Begriff geprägt hat, diesen positiv zu besetzen (im Sinne eines *Reclaim*, einer Rückeroberung negativ besetzter Begriffe), ist, was die Einstellungen der breiten Bevölkerung betrifft, nicht gelungen (vgl. hierzu auch Androutsopoulos 2007).

⁵ So beträgt beispielsweise der Anteil bei den unter 5-Jährigen in Berlin 44%, in Hamburg 47%, in München 53% und in Stuttgart 59% (Quelle: Statistische Ämter, Angaben 2010/11).

Das im Folgenden vorgestellte Korpus hat einen Schwerpunkt auf dem Sprachgebrauch Jugendlicher in einer solchen multiethnischen Population in Berlin-Kreuzberg (Hauptkorpus), umfasst darüber hinaus jedoch auch vergleichbare Daten zum Sprachgebrauch Jugendlicher in einem monoethnischen Wohngebiet, in dem der traditionelle Berliner Dialekt verbreitet ist, nämlich Berlin-Hellersdorf (Ergänzungskorpus). Das KiDKo liefert damit zum einen erstmals eine umfangreiche, spontansprachliche Datenbasis zu Jugendsprache in multiethnischen Wohngebieten und ist damit unseres Wissens einmalig im deutschsprachigen Raum. Zum anderen kann es als empirische Grundlage für weitergehende Untersuchungen zu Jugendsprache ebenso wie zu stadt- und regionalsprachlichen Phänomenen im Berlinischen sowie zum informellen, dialogischen Sprachgebrauch dienen. Es stellt damit eine wertvolle neue Ressource für die Erforschung von Fragen zu Sprachkontakt, Sprachvariation, Sprachgebrauch und Sprachstruktur dar und kann als empirische Basis für grammatische Untersuchungen ebenso wie für solche aus weiteren linguistischen Teildisziplinen dienen, etwa Variationslinguistik, Gesprochene-Sprache- und Gesprächsforschung, (Diskurs-)Pragmatik, Soziolinguistik, DaZ, Sprachkontakt- und Mehrsprachigkeitsforschung. Aus computerlinguistischer Perspektive führt die geplante weitere Aufbereitung des KiDKo zur Weiterentwicklung bestehender Tools durch automatische Verfahren gestützte Aufbereitung spontansprachlicher Korpora (bspw. Tagging und Parsing).

2. Was macht Kiezdeutsch für korpusgestützte linguistische Analysen interessant?

Mit Kiezdeutsch ist im KiDKo eine Varietät repräsentiert, die sich zunächst durch ihr junges Alter und durch die sprachlich heterogene Zusammensetzung ihrer Sprecher/innen auszeichnet. Den Hintergrund für die Herausbildung dieser Sprachpraxis bildet ein multilinguales Setting, das durch vielfältigste linguistische Einflüsse eine hohe sprachliche Dynamik entfaltet. Im Laufe seiner ca. 20-30-jährigen Entwicklung⁶ hat Kiezdeutsch systematische linguistische Strukturen ausgebildet, die es als eigenständige Varietät ausweisen; damit kann Kiezdeutsch heute durchaus als ein neuer urbaner Dialekt begriffen werden. Die Einordnung als Dialekt betont die Herausbildung einer eigenen Dialektgrammatik in Kiezdeutsch, d.h. eines eigenständigen sprachlichen Systems, das nicht durch willkürliche Fehler entsteht, sondern systematische Charakteristika aufweist, die ein Zusammenspiel unterschiedlicher grammatischer Teilsysteme und ihre Interaktion mit außergrammatischen Systemen wie Informationsstruktur und Diskursorganisation widerspiegeln (vgl. Freywald et al. 2011). Ebenso wie andere Dialekte kann Kiezdeutsch in unterschiedlichen Gesprächssituationen und je nach Sprecher/in variabel in stärker oder schwächer ausgeprägter Form auftreten. Im Unterschied zu herkömmlichen Dialekten ist Kiezdeutsch nicht geografisch in einer bestimmten Region verankert, sondern charakteristisch für bestimmte, nämlich mehrsprachig dominierte, Wohngebiete.⁷

⁶ Das Alter von Kiezdeutsch kann man nur schätzen, sprachwissenschaftliche Untersuchungen zu Jugendsprache im multiethnischen Kontext liegen seit Mitte der 1990er Jahren vor (z.B. Tertilt 1996; Auer/Dirim 1999; Füglein 2000; Keim/Androutsopoulos 2000; Keim 2001).

⁷ Dies setzt damit einen weiten, nicht auf regionale Varietäten beschränkten Dialektbegriff voraus, wie er unter anderem in der Variationslinguistik gebräuchlich ist (vgl. etwa Chambers/Trudgill 1998). Für eine ausführliche Diskussion der Einordnung von Kiezdeutsch als Dialekt vgl. Wiese (2012, ersch.).

Charakteristisch für Kiezdeutsch ist das gleichzeitige Vorhandensein von jugendsprachlichen Merkmalen und Eigenschaften einer Kontaktsprache. Viele Sprecher/innen sind mehrsprachig, d.h. sie sprechen neben Deutsch auch Türkisch, Arabisch, Kurdisch, Persisch, Bosnisch, Serbisch u.a.m. Konstitutiv für Kiezdeutsch ist, dass auch Sprecher/innen ohne Migrationshintergrund, also monolingual deutschsprachige Jugendliche, zur Sprechergruppe zählen. Diese Zusammensetzung macht den urbanen Dialekt Kiezdeutsch zu einer Varietät, die unter den Begriff des *Multiethnolekts* subsumiert werden kann (vgl. Quist 2000; Clyne 2000 zum Begriff des Multiethnolekts; Freywald et al. 2011).

Zu Kiezdeutsch und vergleichbaren Sprechweisen liegt inzwischen eine Reihe von Untersuchungen vor, die sich zum Teil vorwiegend soziolinguistischen Aspekten widmen (z.B. Kallmeyer/Keim 2002; Eksner 2006; Androutsopoulos 2007; Keim 2007), zum Teil (auch) die grammatischen Beschreibungsebenen stärker in den Blick nehmen (etwa Füglein 2000; Auer 2003; Dirim/Auer 2004; Kern/Selting 2006; Wiese 2006, 2009). Bezeichnend für diese in verschiedenen Städten Deutschlands durchgeführten Untersuchungen ist, dass die Beschreibungen linguistischer Charakteristika in hohem Maße konvergieren. Offensichtlich liegen hier Entwicklungen vor, die unter denselben Bedingungen entstehen und ähnlich verlaufen und daher möglicherweise universale Tendenzen widerspiegeln. Hierfür sprechen auch die Studienergebnisse zu anderen germanischen Sprachen, in denen sich in den letzten Jahrzehnten bei Sprecher/innen in urbanen multiethnischen Räumen ganz ähnliche Sprechweisen mit vergleichbaren Eigenschaften herausgebildet haben, so etwa im Norwegischen (Svendsen 2010), im Schwedischen (Kotsinas 1992, 1998; Bijvoet 2003; Fraurud 2003; Ganuza 2008), im Dänischen (Quist 2000, 2005, 2008) und im Niederländischen (Appel 1999; Nortier 2001; Cornips 2004).

Als übergreifende Eigenschaft von Kiezdeutsch und seinen europäischen Pendants, die sich unabhängig voneinander entwickelt haben, lässt sich festhalten, dass sich hier Sprachwandelprozesse in einer Situation vollziehen, die von Mehrsprachigkeit und einer hohen Dynamik im Sinne einer größeren Toleranz gegenüber Variation geprägt ist, so dass sich universale linguistische Präferenzen und Entwicklungspfade direkter zeigen als in Sprachen, die durch eine lang(sam)e Entwicklungsgeschichte und durch Normierungsprozesse stärker restringiert sind (vgl. Cheshire et al. 2008, 2011; Wiese 2009, 2011a).

Zur Veranschaulichung dieser Prozesse in Kiezdeutsch führen wir hier exemplarisch drei Phänomene an, die zum einen auf eine verstärkte Verwendung von Nichtstandard-Konstruktionen hinweisen, die auch aus anderen Varianten des Deutschen bekannt sind, nämlich den Gebrauch von *voll* als Intensivierer, und zum anderen auf die Herausbildung neuer grammatischer Muster, nämlich den Gebrauch von monomorphematischem *gibs* (aus existenziellem *gibt es* bzw. *gibt's*) als Existenzpartikel und die Verwendung von bloßen Nominalphrasen (ohne Determinierer und Präposition) als Lokal- und Direktionalangaben. Auf zwei weitere Phänomenbereiche, nämlich auf informationsstrukturell motivierte Verwendungen der Partikel *so* und auf topologische Variation in der linken Satzperipherie gehen wir in Kapitel 4 ausführlich ein.

Die Verwendung von *voll* mit intensivierender Bedeutung ist generell aus gesprochener informeller Sprache und insbesondere auch aus Jugendsprache bekannt (vgl. etwa Androutsopoulos 1998; Kirschbaum 2002; Pittner/Berman 2006). *voll* steht als Intensivierer mit Adjektiv-, Adverb- und Verbphrasen sowie mit Determinansphrasen in Konstruktionen des Typs „*voll* Definitartikel NP“. In Kiezdeutsch finden sich sämtliche dieser Ver-

wendungsweisen. (1) gibt einige Beispiele aus dem multiethnischen Hauptkorpus; (2) stellt dem vergleichbare Daten aus der im Ergänzungskorpus erfassten monoethnischen Sprechergemeinschaft gegenüber:⁸

- | | | |
|------|---|-----------------------------|
| (1a) | oh ich bin heut morgen voll LUSTig , | (KiDKo, Transkript MuH11MD) |
| (1b) | das habt ihr aber WIRKLICH voll oft ne ? | (KiDKo, Transkript MuH1WD) |
| (1c) | isch ZITter immer voll und dann verRUTSCHT es jedes mal . | (KiDKo, Transkript MuH19WT) |
| (1d) | er is voll das KIND . | (KiDKo, Transkript MuH12MD) |
| (1e) | isch bin voll das BÜscherwurm | (KiDKo, Transkript MuH19WT) |
| (2a) | dass sie ihn voll SÜSS findet . | (KiDKo, Transkript Mo01MD) |
| (2b) | musst ick erstma voll LACHen . | (KiDKo, Transkript Mo18MD) |
| (2c) | yeah KUCK ma , voll die GANGster vor uns . | (KiDKo, Transkript Mo05WD) |

Wir finden hier in Kiezdeutsch somit dieselben Muster wie in vergleichbarer Jugendsprache aus stärker einsprachig geprägten Kontexten. Ein Vergleich der Auftretenshäufigkeiten im Korpus weist jedoch darauf hin, dass diese Nichtstandard-Verwendung von *voll* im Gebrauch der multiethnischen Population stärker verbreitet ist als in der monoethnischen Population (vgl. Wiese, ersch.): Sie kommt im Hauptkorpus mehr als doppelt so häufig pro 1.000 Wörter vor wie im Ergänzungskorpus, während *voll* in herkömmlicher Verwendung (= in der Bedeutung „gefüllt“) in beiden Teilkorpora fast gleich oft pro 1.000 Wörter auftritt. Dies weist daraufhin, dass wir in Kiezdeutsch eine Varietät vorfinden, in der sprachliche Entwicklungen in informeller Sprache sehr deutlich werden und die somit für Untersuchungen zu Nichtstandard-Phänomenen besonders interessant ist.

Das Auftreten von *gibs* als Existenzpartikel deutet demgegenüber auf eine Konstruktion, die sich spezifisch in Kiezdeutsch entwickelt. (3) gibt zwei Beispiele aus dem Hauptkorpus:

- | | | |
|------|---|----------------------------|
| (3a) | GUCK ma was hier alles NOCH gibs . | (KiDKo, Transkript MuP1MK) |
| (3b) | WEIßte doch . die die in verschiedene FARBen gibs ? | (KiDKo, Transkript MuH9WT) |

In den subordinierten Sätzen in (3) wird das *-s* in *gibs* offenbar nicht mehr als klitisiertes Subjekt interpretiert, sondern als Teil einer Partikel, die in der Position des finiten Verbs steht; sonst müsste es – als Vollform *es* – in der Subjektposition erscheinen (etwa ..., *was es hier alles noch gibt*). Dies geht möglicherweise einher mit einer Reinterpretation des ursprünglich akkusativischen Objekts (*was* bzw. *die*) als (formgleicher) Nominativ und damit als Subjekt. Eine solche Reinterpretation wird systemintern durch die Syntax-Semantik-Alignierung der Konstruktion gestützt (hierzu ausführlich Wiese/Duda, ersch.): Während in der ursprünglichen Verwendung von *geben* als Existenzverb die Zuordnung thematischer Rollen zu syntaktischen Funktionen aus Sicht des grammatischen Systems deviant ist (das Subjekt *es* erhält keine thematische Rolle und die höchste thematische Rolle, in diesem Fall das Thema, ist mit dem Objekt assoziiert), führt die Reinterpretation zu einer Regularisierung (keine syntaktischen Konstituenten ohne semantisches Pendant; das Subjekt erhält die höchste thematische Rolle). Vorläufer dieser Verwendung von *gib(t)s* sind vermutlich Konstruktionen, in denen ein volles pronominales Subjekt *es* mit *gib(t)s* gemeinsam auftritt, d.h. in denen die Interpretation von *gib(t)s* als Verschmelzung

⁸ Zur Erläuterung der Sprechersiglen vgl. Abschn. 3.4.

von finitem Verb und Subjektpronomen bereits verloren gegangen ist („es gib(t)s ...“, ... gib(t)s es ...“). Diese Verbindungen finden sich in Kiezdeutsch ebenso wie in anderen Varianten des umgangssprachlichen Deutschen (vgl. Wiese 2012; Wiese/Duda, ersch.):

- (4a) isch dachte es gibts nur FÜNFeuroladen. (KiDKo, Transkript MuH9WT)
 (4b) Es gibts nichts schöneres als mit einem guten Buch mit einem Glas Rotwein vor einem offenen Kamin zu sitzen.⁹

Die Beispiele in (5) illustrieren die Verwendung von bloßen Nominalphrasen als Lokal- und Direktionalangaben:

- (5a) un wenn du mal PARty bist, [...] (KiDKo, Transkript MuH11MD)
 (5b) seid ihr macFIT gegang? (KiDKo, Transkript MuH23MT)
 [McFIT = Name eines Fitness-Studios]

Bloße Nominalphrasen als Lokalangaben finden sich auch in anderen informellen Varianten des gesprochenen Deutschen, sind hier jedoch typischerweise auf den Bereich der Haltestellenbezeichnungen beschränkt („Wir sind Alexanderplatz“; vgl. Wiese 2006, 2012). Der Gebrauch in Kiezdeutsch kann somit als eine Generalisierung dieses auch sonst auftretenden Musters verstanden werden.

Wie hier bereits deutlich wird, ist Kiezdeutsch eine Varietät, die durch ihre sprachliche Dynamik besonders interessant ist und dabei Innovationen aufweist, die grundsätzlich im grammatischen System des Deutschen verankert sind und aus diesem heraus motiviert werden können. Untersuchungen zu dieser Varietät sind daher nicht nur aus kontaktsprachlicher Sicht von Bedeutung, sondern gerade auch dazu angetan, ein Licht auf systeminterne Prozesse des Deutschen zu werfen. Das KiDKo stützt dies weiter durch die Verbindung von Daten aus der multiethnischen Sprechergemeinschaft (Hauptkorpus) mit solchen aus einer vergleichbaren monoethnischen Sprechergemeinschaft (Ergänzungskorpus).¹⁰ Es gibt damit ein Mittel an die Hand zur Identifikation und Analyse real gesprochener urbaner Jugendsprache und zur Untersuchung allgemeiner gesprochensprachlicher Phänomene des Deutschen.

Das KiDKo gestattet dabei nicht nur, mit reliablen, authentischen Daten zu arbeiten, sondern erlaubt auch quantitative Untersuchungen, wie etwa Frequenz- oder Overuse-/Underuse-Analysen, die sehr viel subtilere Aussagen über das grammatische System und über Sprachwandelprozesse zulassen. Darüber hinaus ermöglicht die Datenmenge des KiDKo (im Gegensatz etwa zu teilnehmender Beobachtung oder Interviews), mit einer Gesamtgröße von ca. 333.000 Token, auch das Auffinden und Systematisieren seltenerer Phänomene.

3. Korpusdesign und Korpuserstellung

3.1 Sprachdaten

Das KiezDeutsch-Korpus enthält Gespräche zwischen Jugendlichen im Alter von 14 bis 17 Jahren, die diese in ihrer Freizeit mit Freund/inn/en, also innerhalb ihrer *Peer-Group*, geführt haben. Die Datengewinnung fand im Juni 2008 in Form von Selbstaufnahmen in

⁹ Wiese (2011b, S. 1002); aus einem Internetforum zu IT News, <http://forum.golem.de/read.php?27931,1463768,1464287> (letzter Zugriff 9.5.2011).

¹⁰ Siehe Abschnitt 3.1 zur Korpuszusammensetzung.

Abwesenheit von Erwachsenen oder gruppenfremden Personen, wie etwa Exploratoren, statt. Bei den im KiDKo versammelten Daten handelt es sich also um spontane, mündliche, informelle, nicht-monologische Sprache. 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.

Der Erstkontakt zu den Jugendlichen wurde über zwei Schulen hergestellt: eine Schule in Berlin-Kreuzberg mit einem Anteil von Schüler/inn/en nicht-deutscher Herkunftssprache von 84,4% und eine Schule in Berlin-Hellersdorf mit einem Anteil von Schüler/inn/en nicht-deutscher Herkunftssprache von 4,8%.¹¹

Das KiDKo besteht zurzeit aus zwei Teilkorpora, die sich hinsichtlich der Sprechergruppen und Erhebungsorte unterscheiden:

- (i) Das *Hauptkorpus* umfasst Aufnahmen von 17 Jugendlichen unterschiedlicher (einschließlich deutscher) Herkunft aus der Kreuzberger Schule;
- (ii) das *Ergänzungskorpus* versammelt Aufnahmen von 6 Jugendlichen deutscher Herkunft aus der Hellersdorfer Schule.

Die beiden Erhebungsgebiete unterscheiden sich also deutlich in Bezug auf die ethnischen Hintergründe innerhalb der Sprechergemeinschaft. Hinsichtlich weiterer sozioökonomischer Daten differieren sie dabei nur minimal. Tabelle 1 gibt die Daten für die weitere Umgebung der beiden Schulen wieder (dies entspricht grob den schulischen Einzugsbereichen).

	Arbeitslose	Arbeitslose 18-25 Jahre	Langzeitarbeitslose	Nicht arbeitslose Empfänger von Existenzsicherungsleistungen	Nicht erwerbstätige Empfänger von Existenzsicherungsleistungen < 15 Jahren
multiethnisches Gebiet (Kreuzberg)	14,8	10,8	6,0	25,2	59,4
monoethnisches Gebiet (Hellersdorf)	14,3	10,2	6,9	13,3	41,7

Tab. 1: Sozioökonomische Daten für die beiden Rekrutierungsorte in Kreuzberg und Hellersdorf (in Prozent)¹²

Der Gesamtumfang des KiDKo beträgt rund 66 Aufnahmestunden, das umfangreichere Hauptkorpus enthält 48 Aufnahmestunden (ca. 228.000 Token), das kleinere Ergänzungskorpus umfasst 18 Aufnahmestunden (ca. 105.000 Token).

3.2 Anker-Sprecher/innen

Bei der Auswahl der Jugendlichen für die Eigenaufnahmen unter Freund/inn/en wurden Zugehörigkeit zur selben Altersgruppe sowie eine möglichst ausgeglichene Geschlechterverteilung angestrebt. Zudem wurde – im Falle des Hauptkorpus – darauf geachtet, dass unterschiedliche sprachliche Hintergründe vertreten sind und dass die Sprecher/innen mit Kiezdeutsch vertraut sind, so dass sie innerhalb der Sprechergemeinschaft besonders gut als *Anker-Sprecher/innen* fungieren können. Hierzu wurden Daten aus einer Evaluationsstudie (Wiese/Freywald/Mayr 2009) genutzt, in der Bewertungen auditiver Stimuli, die

¹¹ Quelle: Berliner Senatsverwaltung für Bildung, Wissenschaft und Forschung.

¹² Quelle: Berliner Senatsverwaltung für Stadtentwicklung, *Stadtmonitoring Soziale Stadtentwicklung*.

u.a. typische Kiezdeutsch-Merkmale enthielten, erhoben worden waren. Alle Anker-Sprecher/innen sind in Deutschland geboren und aufgewachsen. In einem Fragebogen wurden Angaben zu Alter, Geschlecht und Wohnviertel erhoben sowie Informationen dazu, welche Sprachen typischerweise in welchem Kontext gesprochen werden mit Eltern, mit Geschwistern, mit Freund/inn/en. Die Merkmale der im KiDKo vertretenen Anker-Sprecher/innen sind in Tabelle 2 zusammengefasst. „Alltagssprache“ gibt hierbei die Sprache wieder, die mit Freund/inn/en vorwiegend verwendet wird, „Familiensprache“ diejenige, die mit den Eltern gesprochen wird (mit Geschwistern werden meist sowohl Alltags- als auch Familiensprache gesprochen, sofern es sich dabei um verschiedene Sprachen handelt).

Teilkorpora Metakategorien	Hauptkorpus (17 Anker-Sprecher/innen)	Ergänzungskorpus (6 Anker-Sprecher/innen)
Alter	14-17 Jahre	15-16 Jahre
Geschlecht	weiblich: 7 männlich: 10	weiblich: 1 männlich: 5
Wohnviertel	Kreuzberg: 12 Neukölln: 2 Mitte: 1 Wedding: 1 Tiergarten: 1	Hellersdorf-Marzahn: 6
Alltagssprache	Deutsch: 17	Deutsch: 6
Familiensprache	Türkisch: 8 Kurdisch: 3 Arabisch: 2 Deutsch: 4	Deutsch: 6

Tab. 2: Erhobene Merkmale der Anker-Sprecher/innen

Die ausgewählten Jugendlichen wurden für die Eigenaufnahmen mit digitalen Aufnahmegegeräten und Ansteckmikrofonen ausgestattet und instruiert, eine Woche lang Freizeitgespräche aufzunehmen, wobei die beteiligten Gesprächspartner/innen jedes Mal vorab über die Aufnahme informiert werden sollten. Über Gesprächspartner/innen, Gesprächsthema und Aufnahmesituation sollte ein Protokoll geführt werden.

3.3 Transkription

Derzeit liegen 66 Stunden Aufnahmen mit Transkriptionen im XML-Format vor. Die Transkriptionen sind mit den Audiodateien aligniert. Um eine möglichst mühelose Lesbarkeit der Transkripte zu gewährleisten und den Transkriptionsaufwand nicht zu stark zu erhöhen, wurde keine phonetische Transkription gewählt, sondern eine, die an die Konventionen des Gesprächsanalytischen Transkriptionssystems (GAT) für das Basistranskript angelehnt ist (vgl. Selting et al. 1998 sowie die aktualisierte Fassung Selting et al. 2009). Damit liegen die Daten in einer literarischen Transkription vor, die weitgehend orthografisch ist, aber dennoch die Notation sowohl von gesprochensprachlichen, varietätsspezifischen und individuellen Besonderheiten als auch von prosodischer Information erlaubt. Die Kennzeichnung prosodischer Merkmale erfolgt dabei in den Transkripten lediglich grob und wurde im Vergleich zum GAT-Basistranskript vereinfacht. Markiert sind Hauptakzente (durch Versalien) sowie Pausen (mit den Werten „kurz“ und „lang“ sowie bei Pausen über drei Sekunden mit einer genauen Zeitangabe).

Bei der Verschriftung gesprochener Sprache ergeben sich naturgemäß zahlreiche Probleme. An erster Stelle steht die Schwierigkeit, mündliche Äußerungen in ‚schriftgemäße‘ Einheiten zu segmentieren. Für die technische Verarbeitbarkeit der Transkripte (wie etwa automatische Suchanfragen) ist es notwendig, Wort- und Satzgrenzen zu markieren. Bekanntermaßen widersetzt sich gesprochene Sprache häufig ebenjener Unterteilung bzw. kann man in vielen Fällen nicht von einer Eins-zu-Eins-Entsprechung dieser Einheiten in geschriebener und gesprochener Sprache ausgehen (vgl. zum Problem von Segmentierungseinheiten in gesprochener Sprache u.a. Foster/Tonkyn/Wigglesworth 2000; Fiehler 2003 sowie Auer 2010).

Auf *Wortebene* treten z.B. folgende Konflikte zwischen gesprochen- und geschriebensprachlichem Wortbegriff auf:

(i) Verschmelzungen

Hierzu zählen durch einen schnellen Anschluss kontrahierte bzw. klitierte Formen, die schriftlich durch zwei (oder mehrere) Einzelwörter wiedergegeben werden, wie z.B. *sa=mal* für *sag mal*, *willst=e* für *willst du*, *is=n* für *ist denn* usw.

(ii) Reduktionsformen

In den seltensten Fällen wird in Explizitlautung gesprochen, welche in etwa durch die orthografische Form wiedergegeben werden könnte, sondern in aller Regel fallen Schwa-haltige Silben aus, silbenfinale Konsonanten werden nicht artikuliert, Funktionswörter werden nicht in ihrer Vollform gebraucht, es finden Assimilationsprozesse statt usw. Solche Aussprachemerkmale sind an der orthografischen Wortform nicht erkennbar. Beispiele aus dem KiDKo sind: *am bestn* für *am besten*, *jetz* für *jetzt*, *glei* für *gleich*, *nomal* für *nochmal*, *ne/nen* für *eine/einen*, *ham* für *haben* u.a.

(iii) Wörter ohne orthografische Entsprechung

Im gesprochenen Deutsch wird eine Reihe von Wörtern verwendet, die in der Schriftsprache nicht vorkommen und die dementsprechend auch keine kodifizierte orthografische Form besitzen. Dies betrifft bestimmte Interjektionen (wie *oha*, *ah ja*, *na na*), Question-tags (*ne*, *wa*), Rezeptionssignale (*hm*, *'hm'hm* (mit Glottalverschlüssen), *nö*, *hā*) und gefüllte Pausen/Verzögerungspartikeln (*äh*, *ähm*). Daneben kommen dialektale Ausdrücke und Aussprachevarianten (in unserem Fall berlinische und kiezdeutschspezifische Wörter) vor, die ebenfalls keine orthografische Entsprechung haben, da sie aus rein mündlichen Varietäten stammen, deren schriftliche Form nicht normiert ist. Hierzu zählen Formen wie *weeßte* für *weißte*, *gloob* für *glaub*, *ölf* für *elf*, *Tach* für *Tag*, *ejal* für *egal*, *wa* für *wir*, *nisch* für *nicht*, *leicht* für *leicht*, *bu* für *zu*, die Interjektionen *wallah* und *çüş* usw.

Auf *Satzebene* besteht das größte Problem bei der Verschriftung gesprochener Sprache darin, Satzgrenzen festzulegen, die den schriftsprachlichen Satzmustern entsprechen. Diese Schwierigkeit ergibt sich direkt aus den Wesensmerkmalen mündlicher Kommunikation, nämlich der zeitlichen Linearität und der Interaktionalität. Bedingt durch den linearen und flüchtigen Verlauf mündlicher Äußerungen wirken sich Planungsprozesse unter Umständen unmittelbar auf die syntaktische Oberflächenstruktur aus. Daneben bedingt bzw. erfordert die Interaktion mit einem/einer oder mehreren Gesprächsteilnehmer/innen spezielle Kommunikationsstrategien, die sich ebenfalls in der syntaktischen Form von Äußerungen widerspiegeln können (vgl. Auer 2007, 2010). Innerhalb des KiDKo finden

sich insbesondere folgende Merkmale spontanen Sprechens, die eine Festlegung von Satzgrenzen (im schriftsprachlichen Sinne) erschweren oder gar unmöglich machen:

(i) Satzabbrüche

Als Satzabbrüche sollen hier nur tatsächlich nicht beendete satzartige Strukturen verstanden werden (zu Umplanung/Umkonstruktion s. (iii) unten). Abbrüche kommen im KiDKo – wie allgemein in informeller Spontansprache – relativ häufig vor. Sie kommen vor allem durch Unterbrechungen von außen zustande, die den Sprecher/die Sprecherin daran hindern, seinen/ihren Redebeitrag zu Ende zu führen. Dies geschieht meist durch Einwürfe anderer Gesprächsteilnehmer/innen, aber auch durch laute Umgebungsgeräusche oder ähnliche Ablenkungen, die die Aufmerksamkeit der Sprecher/innen anderweitig in Anspruch nehmen.

(ii) Setzungen

Eine weitere Kategorie von Syntagmen, die einer schriftsprachlich orientierten Satzdefinition kaum genügen, bilden syntaktisch relativ autonome, selbst nicht satzförmige Äußerungseinheiten, die wir möglichst neutral mit dem Begriff ‚Setzungen‘ bezeichnen wollen. Gemeint sind damit nicht Ellipsen im engeren Sinn oder so genannte Satzäquivalente, vielmehr sind hierzu Einzelwörter und -phrasen zu rechnen, die außerhalb des eigentlichen Satzverbandes stehen und auch nicht als eine Form von Vorvorfeld- oder Nach-/Schlussfeldbesetzung interpretiert werden können, vgl. die Beispiele in (6). (In den Transkriptionen sind Personennamen im Format ‚Anfangsbuchstabe + xxx‘ anonymisiert (z.B. pxxx für *Peter*) bzw., wenn kein Name bekannt ist, als ‚spk1‘, ‚spk2‘ usw. (spk = speaker); bei Akzentuierung erscheint das Kürzel in Versalien, wie z.B. IXXX in (6) oder bspw. SPK1. Die Sprecherebenen sind einheitlich nach dem Muster ‚großer Anfangsbuchstabe + xxx‘ für anonymisierte Sprecher und ‚Spk1, Spk2‘ etc. bei unbekanntem Namen benannt.) Als diskursstrukturierende Elemente (wie *eGAL*), Reklamationen des Rederechts (*auf jeden FALL*), evaluative Kommentare (*voll LANGweilig*), Nachträge (*spiel, türkei gegen DINGS, schweiz; dings gegen DINGS*) usw. weisen sie lediglich einen inhaltlichen Bezug zur vorherigen oder folgenden Äußerung bzw. zur aktuellen Sprechsituation auf.

(6) Verabredung zum Fußballschauen während der WM

Exxx: heißt das MILCHshake ?
 MuH17MA: was wir gestern geHOLT haben du SCHWUCHtel .
 Exxx: ja :.
 MuH17MA: bei IXXX .
 Exxx: MILCHshake ?
 MuH17MA: ja: MILCHshake . ähm gehn wa nachher DINGS gucken ?
 spiel ? türkei gegen DINGS ? schweiz ?
 [...]
 MuH17MA: eGAL exxx . auf jeden FALL exxx , (-) gehn wir nachher SPIEL gucken ? dings gegen DINGS ?
 [...]
 MuH17MA: voll LANGweilig was MACHen wir ?
 Exxx: lassma noch CHILLN ,
 (KiDKo, Transkript MuH17MA)

(iii) Umplanungen

Strategien der Redebeitragskonzipierung und ähnliche Planungsprozesse schlagen sich häufig direkt in der Äußerungsstruktur nieder, d.h. einmal geäußerte Phrasen oder Syntagmen werden nicht erwartungsgemäß zu Ende geführt, sondern offen gelassen oder aber als Bestandteile einer neuen syntaktischen Struktur uminterpretiert. Es kommt zu Selbstreparaturen, Umkonstruktionen und Neueinsätzen, wie etwa in (7):

(7) Gespräch über fünf Euro Taschengeld pro Woche

MuH17MA: fünf EUro , auch wenn du NICH rauchen würdest würde nisch reichen fünf euro .
 Exxx: ja: . (-) na GU= ma , IS so: . **ich (-) du WEIBT dass ich JEden tag , (-)** weil ich RAUche muss ich jeden tag RAUchen . **also (-) brauch ich eigentlich VIER euro für , (-)** (dis) VIER euro muss standard sein für DINGS . (-) SCHACHtel .

(KiDKo, Transkript MuH17MA)

Diese typischen Charakteristika gesprochener Sprache stellen große konzeptuelle und technische Herausforderungen an die weitere Aufbereitung des Korpus, insbesondere an das syntaktische Parsen der Daten. Dies wird Gegenstand der nächsten Arbeitsschritte sein (vgl. Kapitel 5).

Neben den grundsätzlichen Schwierigkeiten, gesprochene Sprache schriftlich abzubilden, hat das KiDKo die Besonderheit, dass es stellenweise Gesprächsanteile in anderen Sprachen als dem Deutschen (z.B. Türkisch, Arabisch) enthält. Dieser Anteil liegt bei ca. 10%. Typischerweise treten solche Äußerungen im Kontext von Code-Switching auf, wie in Beispiel (8). Dass ganze Gesprächssequenzen nicht deutschsprachig sind, ist eher die Ausnahme. In (8) wechseln beide Sprecherinnen zwischen Deutsch und Türkisch, vgl. die Rückfrage *ne* ‚was‘ sowie die Aufforderung *bunun içine koy* ‚tu das hier rein‘ (wörtl.: ‚dieses hinein tu‘):

(8) Gespräch zwischen zwei Freundinnen zu Hause

Spk1: wann stehen wir morgen AUF ?
 MuH27WT: ne ?
 Spk1: wann stehen wir morgen AUF ?
 MuH27WT: äh u::m früh MORgens .
 Spk1: punkt SIEben .
 [...]
 Spk1: GXXX . diesen haar WIRFST du nich auf =n boden .
 MuH27WT: (unverständlich)
 Spk1: **bunun içine koy** .
 MuH27WT: hm .
 (KiDKo, Transkript MuH27WT)

3.4 Korpusaufbau in EXMARaLDA

Für Transkription und (Teil-)Annotation wurden die im SFB 538 „Mehrsprachigkeit“ der Universität Hamburg entwickelten Tools des Programms EXMARaLDA (Extensible Markups Language for Discourse Annotation) verwendet, ein System zur computergestützten Erstellung und Analyse von Korpora gesprochener Sprache (vgl. Schmidt/Wörner 2005). Die grafische Oberfläche des Partitur-Editors, auf dessen Basis die Transkriptionen angefertigt wurden, stellt eine Transkription als potenziell unbegrenzt lange Partitur dar,

bestehend aus einzelnen Events, in diesem Fall Wörtern.¹³ Da es sich um dialogische Daten handelt, sind die einzelnen Sprecher/innen in Form unterschiedlicher Spuren repräsentiert, die es erlauben, Sprecherwechsel sowie Überlagerungen beim Sprechen wortgenau abzubilden.

Zu jedem Transkript sind Metadaten zu den Anker-Sprecher/innen sowie zur Gesprächssituation in der XML-Datei gespeichert. Dazu gehören die oben in Tabelle 2 zusammengefassten Metakategorien sowie Meta-Informationen zum Gespräch als Ganzem (bspw. Ort des Gesprächs) sowie zur Aufnahme (bspw. deren Qualität, mögliche Störgeräusche).

Die Aufnahme Dateien wurden jeweils unter der Sigle des Anker-Sprechers/der Anker-Sprecherin erfasst. Die Siglen sind „sprechende“ Kürzel, d.h. sie geben Auskunft zu bestimmten Eigenschaften der betreffenden Sprecher/innen. Eine Sigle, etwa MuH9WT oder Mo18MD, setzt sich aus vier Bestandteilen zusammen: Mu|H9|W|T. Dabei steht die erste Variable (mit den Werten Mu und Mo) für multiethnisches bzw. monoethnisches Wohngebiet, die zweite Variable enthält die fortlaufende Sprechernummer¹⁴, die dritte Variable bezeichnet das Geschlecht (mit den Werten W und M für weiblich/männlich) und die vierte Variable gibt Auskunft über die Familiensprache (mit den Werten T, K, A und D für Türkisch, Kurdisch, Arabisch, Deutsch). Hinter MuH9WT steht also eine Sprecherin mit Türkisch als Familiensprache, die aus dem multiethnischen Wohngebiet Kreuzberg stammt, Mo18MD bezeichnet einen männlichen, aus dem monoethnischen Hellersdorf stammenden Sprecher mit Deutsch als Familiensprache.

Die Transkriptionen sind mit den entsprechenden Audiodateien verlinkt und zurzeit in Abständen von etwa zwei Minuten¹⁵ mit diesen aligniert, so dass jede Belegstelle im Transkript an die ursprüngliche Audiosequenz rückgebunden ist. Abbildung 1 zeigt den Ausschnitt einer EXMARaLDA-Partitur aus dem KiezDeutsch-Korpus. Es handelt sich dabei um den Ausschnitt eines Anker-Sprecher-Transkripts mit der Sigle MuP1MK. Neben den Sprachdaten des Anker-Sprechers enthält dieses Transkript Daten des Gesprächspartners Spk1.

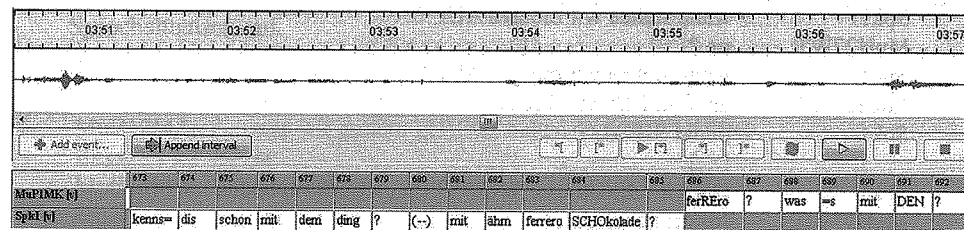


Abb. 1: Ausschnitt einer EXMARaLDA-Partitur aus dem KiezDeutsch-Korpus

Aus den einzelnen Transkripten und deren Metadaten kann ein zusammenhängendes Korpus mit Hilfe des Corpus-Managers (CoMa) erstellt werden, der ebenfalls ein Teil des EXMARaLDA-Systems ist. In Abhängigkeit von der Fragestellung ist mit CoMa sowohl

¹³ Die Transkriptionen wurden mit dem GENERIC-Algorithmus segmentiert. Ein Token ist dabei definiert als eine Zeichenkette, die entweder durch Leerzeichen oder Interpunktion begrenzt ist.

¹⁴ Der Zusatz „H“ bzw. „P“ verweist auf unterschiedliche Durchläufe der erwähnten Evaluationsstudie (Wiese/Freywald/Mayr 2009), an der die Sprecher/innen teilgenommen haben.

¹⁵ Alignierungsgrenzen können durch Störgeräusche in den Tondateien minimal abweichen.

die Erstellung eines Gesamtkorpus aller Transkriptionen möglich als auch die Klassifizierung einzelner Subkorpora, bspw. mit Bezug auf bestimmte Merkmale in den Metadaten der Transkriptionen (z.B. Transkripte von Sprecher/innen einer bestimmten Familiensprache oder eines bestimmten Alters).

Für die Extraktion von Belegen zu bestimmten Phänomenen und deren quantitative Analyse steht in EXMARaLDA das Analyse- und Konkordanztool EXAKT zur Verfügung. Mit EXAKT sind gezielte Suchanfragen an die Transkriptionsdateien möglich. Diese Suchanfragen können sich auf transkribiertes verbales Material, auf die Beschreibung von non-verbalem Material oder aber auf Annotationsebenen beziehen. In Zukunft ist die Implementierung in die ANNIS-Mehrebenen-Architektur (Zeldes et al. 2009) geplant, s. dazu Kapitel 5.

4. Linguistische Analysen im Korpus

Das Korpus bietet eine umfangreiche empirische Grundlage für quantitativ und qualitativ basierte Analysen spontansprachlicher Phänomene im multiethnischen und monoethnischen urbanen Raum. Zurzeit sind mit Hilfe von EXAKT bereits Gesamterhebungen für lexikalische Elemente oder feste Wendungen möglich, für die Anfragen auf der Wortebene durchgeführt werden können. Untersuchungen zu syntaktisch komplexen Phänomenen können in der gegenwärtigen Fassung des Korpus noch nicht den gesamten Datenbereich über automatische Suchverfahren einbeziehen, sondern müssen auf Stichproben und/oder auf manuellen Suchen basieren. Die geplante weitere Aufbereitung des Korpus (s.u., Kapitel 5) wird unter anderem Annotationen liefern, die Wortarten, topologische Felder und Phrasengrenzen erfassen und damit umfassende, automatische Suchanfragen auch zu komplexen Phänomenen erlauben.

Im Folgenden stellen wir zwei Beispiele für linguistische Analysen im Korpus vor, die bereits in der gegenwärtigen Fassung durchgeführt werden konnten und die das Potenzial des Korpus illustrieren: (1) eine Untersuchung zu *so* als Fokusmarker, die auf einer Gesamterhebung von *so*-Vorkommnissen basiert, und (2) eine Untersuchung zur linken Satzperipherie, die eine manuell gezogene Stichprobe mit einer Gesamterhebung zu ausgewählten satzinitialen Adverbialen kombiniert.

Die Untersuchungen bezogen neben dem Hauptkorpus kontrastiv das Ergänzungskorpus ein und konnten so mögliche Unterschiede zwischen dem Sprachgebrauch in der multiethnischen und der monoethnischen Population prüfen. Wie im Folgenden deutlich wird, illustrieren die Ergebnisse zu den beiden Phänomenbereichen dabei unterschiedliche Typen von Befunden aus solchen Vergleichen: Bezogen auf das Korpus finden wir Phänomene, die nur im multiethnischen Teil des Korpus auftreten und damit auf *rein* kiezdeutschspezifische Muster verweisen könnten, ebenso wie solche, die in beiden Teilen des Korpus auftreten, im multiethnischen Teil des Korpus möglicherweise etwas häufiger, jedoch grundsätzlich auf ähnlichen Distributionen basieren und damit auf Muster hinweisen, die im System des Deutschen bereits angelegt sind.

4.1 *so* als Fokusmarker

so ist im Deutschen ein multifunktionales Element, das in seiner Kernbedeutung indexikalisch Modalität ausdrückt und in expliziten oder impliziten (und z.T. intensitätsanzeigenden) Vergleichen auf „Wie?“ antwortet (*so groß; so ein Hut wie dieser*). Abgeleitet von dieser Verwendung kann *so* auch als Quotativ-Marker (*Ich so: „Was machst du da?“*;

„Das Maß ist voll.“ so der Vorsitzende) und als Heckenausdruck und/oder Anzeiger für Vagheit (*Das kostet so 50 Euro.*) auftreten.¹⁶ Eine Verwendung, die noch weiter von der indexikalisch-modalen Kernbedeutung entfernt ist, ist die als Fokusmarker (vgl. Wiese 2011b): *so* ist hier semantisch gebleicht und trägt nicht mehr zur lexikalischen Bedeutung des Satzes bei, sondern hat eine rein informationsstrukturelle Funktion, nämlich die Markierung des Fokusausdrucks, der als Ko-Konstituente zu *so* steht. (9) gibt zwei Beispiele (Transkriptionen von uns): (9a) stammt aus einer Fernsehdiskussion (aus dem DWDS-Kernkorpus: „Das literarische Quartett“ vom 14.12.2001, Gesprächsbeitrag von Iris Radisch über „Das erste Jahr“ von Durs Grünbein); (9b) ist ein Hörbeleg aus einem Gespräch zwischen einer Ärztin und ihrer Patientin (Ärztin erklärt die Anwendung einer Creme).

(9a) ich habe das buch NICHT ungern gelesen, weil ich mich sehr amüSIERT habe, nur habe ich den verDACHT, dass das was ich so KOMisch fand, zu GROßen strecken zumindestens gar nicht beABSichtigt war nämlich eine UNfreiwillige komik war. es kommt SEHR gelehrt daher [...] es wird viel erKLÄRT. WER ist nietzsche, waRUM, ich frage mich oft WEM wird das erklärt? es hat so VOLKShochschulhafte züge. das ist aber natürlich noch nicht WAS ich so komisch fand. was ich WIRKlich sehr komisch fand und was doch der domiNANTE ton in diesen tagebüchern ist ist einfach die humaNistische tradition, mit der grünbein ja DURCHweg argumentiert

(9b) die ist für die NACHT und diese so für TAGSüber so.

so tritt als Fokusmarker insbesondere in informellen Varianten des Deutschen, d.h. in gesprochener Sprache, aber auch in Internet-Chats u.ä. auf. In Kiezdeutsch scheint die Verwendung als Fokusmarker besonders ausgeprägt zu sein. Um diesen zunächst vorläufigen Eindruck zu überprüfen, wurde das Auftreten von *so* als Fokusmarker im Korpus quantitativ und qualitativ analysiert und dabei auch für die beiden im Korpus repräsentierten Populationen verglichen. Die Korpusstudie zu *so* fand Ende 2009 statt; Grundlage war der damals verfügbare Korpusumfang von 140.952 Token im Haupt- und 43.939 Token im Ergänzungskorpus.

Im ersten Schritt wurden zunächst im multiethnischen Hauptkorpus alle Vorkommnisse von *so* extrahiert. Mit Hilfe von EXAKT wurde hierfür das gesamte Hauptkorpus auf der Transkriptebene automatisch nach *so* durchsucht; EXAKT lieferte hierfür Konkordanzlisten und die dazugehörigen Partituren, die alle Vorkommnisse von *so* mit ihren sprachlichen Kontexten enthielten. Insgesamt erbrachte die Suche 2.501 Fundstellen für *so*, d.h. rund 18,1 Vorkommnisse pro 1.000 Wörter.

Die Vorkommnisse wurden dann manuell klassifiziert mit dem Ziel, Verwendungen von *so* als Fokusmarker zu identifizieren. Dies wurde von zwei Klassifikator/inn/en unabhängig durchgeführt; Unterschiede in den Klassifikationen und Zweifelsfälle wurden anschließend diskutiert. Die Kriterien für die Einordnung eines *so*-Vorkommnisses als Fokusmarker waren die folgenden: (a) *so* leistet keinen Beitrag zur Satzbedeutung, (b) *so* ist unbetont und (c) *so* tritt mit einer Ko-Konstituente auf, die den Fokus trägt (bzw. den Fokusträger enthält) und daher die Hauptbetonung erhält. Auf dieser Grundlage wurden 560 der *so*-Vorkommnisse im Hauptkorpus übereinstimmend als Fokusmarker-Verwendungen klassifiziert; das sind rund 4,0 Vorkommnisse pro 1.000 Wörter. (10) gibt einige Beispiele:

¹⁶ Zur Diskussion verschiedener Verwendungsweisen von *so* vgl. etwa Ehlich (1987), Golato (2000), Thurmair (2001), Weinrich (2003), Auer (2006), Wiese (2011b).

- (10a) *Gespräch über das Verhalten eines Freundes während der Fußball-WM*
er is ENGLänder und er feiert MIT uns. er hat so türKEItrikot und türkeiFAHne um sich (KiDKo, Transkript MuH13MT)
- (10b) *Gespräch über einen Film, der gerade im Fernsehen läuft*
das sieht so INdisch aus (KiDKo, Transkript MuH1WD)
- (10c) *Diskussion eines Fußballspiels*
aBU, war voll mies. er wird so geFOULT (KiDKo, Transkript MuH12MD)
(abu: Ausdruck der Empörung, arabischen Ursprungs; vgl. Wiese 2012)
- (10d) *Gespräch über Fußballspiel*
aber rumänien hat rischtisch geil verTEIdigt so (KiDKo, Transkript MuH3WT)
- (10e) *Gespräch beim Essen in „Burger King“*
JA: die guckt so zu DIR so (KiDKo, Transkript MuH12MD)

Wie die Belege illustrieren, erfährt *so* als Fokusmarker eine syntaktische Ausweitung und tritt mit allen lexikalischen Hauptkategorien bzw. ihren Expansionen auf, nämlich mit NP/DP (10a), AP (10b), VP (10c,d) und PP (10e). *so* zeigt dabei unterschiedliche Linearisierungsoptionen; es kann in prä-, post- und zirkum-Position auftreten (prä-Position: (10a,b,c); post-Position: (10d); zirkum-Position: (10e)).

Die Suche im monoethnischen Ergänzungskorpus ergab ebenfalls Fundstellen für *so* in der Funktion eines Fokusmarkers. Wie (11) illustriert, fanden wir hier dieselben Distributionsoptionen wie im multiethnischen Hauptkorpus:

- (11a) *Gespräch über das Aufnahmegerät, das für die Aufnahmen genutzt wird*
Spk6: wieviel hast =n beZAHLT dafür.
Mo01MD: NICHTS. dis KRIEG ich hab ich so KOSTenlos bekomme (KiDKo, Transkript Mo01MD) [prä-so, mit AP]
- (11b) *Beginn einer Erzählung über Aktivitäten am Vortag*
ey PXXX. gestern war ich so am biesdorfer SEE we:ßte? hab me: da mit n paar KUMpel getroffen (KiDKo, Transkript Mo18MD) [prä-so, mit PP]
- (11c) *Gespräch über eine Schlägerei*
irgendwie hat er dich so geSCHLAGen und du bist irgendwie kurz SO wa? (KiDKo, Transkript Mo01MD) [prä-so, mit VP]
- (11d) *Beschreibung einer jungen Frau*
übelste ACHselhaare so. (-) MANNsweib. (KiDKo, Transkript Mo18MD) [post-so, mit NP/DP]
- (11e) *Diskussion von Kleidungsstilen auf Hochzeiten*
ma jetze wenn du zu ANderen hochzeiten gehst. so ganz SCHWARZ so. (KiDKo, Transkript Mo05WD) [zirkum-so, mit AP]

Im Ergänzungskorpus (mo) kommt *so* als Token etwa gleich oft pro 1.000 Wörter vor wie im Hauptkorpus (mu), nämlich rund 17,7 Mal (im Vergleich zu 18,1 Mal im Hauptkorpus, s.o.). Als Fokusmarker tritt *so* im Ergänzungskorpus etwas seltener auf, nämlich 145 Mal; das sind 3,3 Vorkommnisse pro 1.000 Wörtern gegenüber 4,0 pro 1.000 Wörtern im Hauptkorpus. Dies könnte auf eine stärkere Etablierung von *so* als Fokusmarker in Kiezdeutsch weisen.

Innerhalb der Fokusmarker-Verwendungen unterscheiden sich die Häufigkeiten der verschiedenen Linearisierungsmöglichkeiten des Fokusmarkers *so* gegenüber seiner Ko-Konstituente nicht signifikant in den beiden Teilkorpora ($\chi^2 = 0,6$; $p = 0,74$); dies weist auf eine systematische Verteilung hin. Die prä-Position ist quantitativ dominant, sie tritt in über 3/4 der Fälle auf (mu: 81,4 % der Fokusmarker-Vorkommnisse von *so*, mo: 78,6 %), gefolgt von der post-Position in etwa 1/7 der Fälle (mu: 13,6 %, mo: 15,9 %), während die zirkum-Position nur in rund 1/20 der Fälle gewählt wird (mu: 5,0 %, mo: 5,5 %).

Wir nehmen an, dass prä-*so* dominant ist wegen der prä-Position von *so* in lexikalischen Verwendungen als indexikalisch-modales Element. Die post-Stellung von *so* könnte gestützt werden durch Verwendungen von *so* am Satzende als Vergewisserungssignal oder, in der Kombination *und so* / *oder so*, als Heckenausdruck, vgl.:

- (12) *Gespräch über eine junge Frau, die gerade Lehrerin geworden ist*
und ähm (-) dann SAGT sie so fragen die SCHÜler so, (-) die MÄdel sie so. ja frau KXXX können sie NICH abends mit uns noch irgendwie WEGgehen **und so**, keine AHnung was, (KiDKo, Transkript MuH1WD)

In der Mehrzahl der Fälle steht fokusmarkierendes post-*so* in der rechten Peripherie des Satzes, und auch hier unterscheiden sich die beiden Teilkorpora nicht signifikant (mu: 57,9 %, mo: 60,9 % der post-*so* stehen am Satzende; $\chi^2 = 0,064$; $p = 0,80$).

Für die post-Stellung von *so* könnte man auch einen Einfluss des Türkischen vermuten, da im Türkischen post-Positionen für Präpositionen und Partikeln dominant sind. Auch wenn es keine Fokusmarker i.e.S. gibt, so hat doch zumindest die Interrogativpartikel *ml*, die ebenfalls in post-Position steht, einen fokusanzeigenden Effekt in Konstituentenfragen, indem sie immer dem Element folgt, nach dem gefragt wird. Einen solchen Einfluss halten wir jedoch nicht für plausibel angesichts unserer Befunde für das Ergänzungskorpus, die Evidenz für dieselbe Verteilung der Stellungstypen in einer monoethnischen Population ohne substanziellen Sprachkontakt mit dem Türkischen liefern.

Einige Belege weisen darauf hin, dass post-*so* der Refokussierung von Common-Ground-Material dienen kann, d.h. dem nochmaligen Verweisen auf bereits Gesagtes, während prä-*so* eher als Default-Fall die neue Information markiert, die dem Common Ground hinzugefügt wird. In (13) markiert das prä-*so* in *so Lila* als Antwort auf die Frage *wie siehts AUS?* die neue Information, während das post-*so* in *nein*, *in Lila so* korrektiven Fokus auf *in Lila* anzeigt und diese Phrase refokussiert:

- (13) *Gespräch über einen Einkauf*
MuH9WT: wir GUCKen so, da is so ein oberteil ja rischtisch SCHÖN,
Spk1: wie siehts AUS?
MuH9WT: **so Lila**, aber GLITZern, weißt doch so voll SCHÖN und (-)
Spk1: war war dis auf TRÄger?
MuH9WT: nein. (-) dis war so TI:shirt und dann noch hier so OFFen, weißtu, so LOCKer.
Spk1: hier OFFen?
MuH9WT: ja.
Spk1: in SILB so GRAU?
MuH9WT: nein, **in Lila so**.
Spk1: isch habs in GRAU geholt von vera MOda, drei EUro wa?
(KiDKo, Transkript MuH9WT)

In zirkum-Position grenzt *so* den Fokusausdruck zu beiden Seiten eindeutig ab und könnte damit einer stärkeren Deliminierung der Fokusdomäne dienen. Der Fokusausdruck wird zudem durch die Deakzentuierung von *so* phonologisch noch stärker hervorgehoben.¹⁷ Die zirkum-Stellung wird möglicherweise durch eine Kombination von post-*so* mit *so* in Progressiv-Verwendung noch weiter gestützt (z.B. *Ich gehe so die Straße lang, da kommt plötzlich ...*). *so* steht in Progressiv-Verwendung am linken Rand der VP, während der Fokusausdruck häufig innerhalb der VP steht; zusammen mit einem post-*so* in rechtsperipherer Position kann dies daher die Entwicklung von zirkum-*so* motivieren.¹⁸ (14a) gibt einen Korpus-Beleg für *so* als Progressivmarker, (14b) illustriert eine Verwendung, die eine Analyse als Kombination aus Progressiv-*so* und post-*so* nahelegt:

- (14a) *Gespräch über ein Foto, auf dem die Sprecherin abgebildet ist*
sie kommt **so** zu mir, sagt so, äh WER ist das [...]
isch guck sie **so** AN, (-)
ich dachte so, dis=s ne DEUTsche,
isch guck so NEE, ne? dis bin ISCH ja.
(KiDKo, Transkript MuH3WT)
- (14b) *Schilderung eines Ereignisses beim Fußballschauen im TV*
wisst ihr was PASSiert ist gestern? (-) ich sitz **so** auf=m Sofa **so**, (-) es kommt ä: (-) die männer schrein so TOR, ich spring aus=m Sofa und SCHREIe.
(KiDKo, Transkript MuH2WT)

Die Belege für die verschiedenen Linearisierungsmöglichkeiten legen damit z.T. zusätzliche funktionale Aspekte für die jeweilige Verwendung von *so* als Fokusmarker nahe (einfache Hinzufügung vs. Refokussierung von Common-Ground-Material; eindeutige Deliminierung der Fokusdomäne). Auch wenn dies zurzeit eher auf Tendenzen als auf strikte Regularitäten hinweist, könnte der jetzige Zustand doch zu einer Stabilisierung und der späteren Etablierung solcher oder weiterer funktionaler Differenzierungen führen.

Zusammengenommen weist die Korpusstudie zu *so* auf ein systematisches Phänomen in der gesprochenen Sprache: *so* wird als Fokusmarker über multi- und monoethnische Populationen hinweg gebraucht und zeigt hier keine qualitativen Unterschiede. Dieser Fokusmarker tritt jeweils mit Ko-Konstituenten derselben syntaktischen Kategorien auf und realisiert dieselben Stellungsmöglichkeiten in Bezug auf seine Ko-Konstituente, ohne signifikante Unterschiede in der Verteilung der Linearisierungen. Der leichte quantitative Vorteil der Verwendung von *so* als Fokusmarker im multiethnischen Hauptkorpus (pro 1.000 Wörter) könnte dabei auf eine mögliche stärkere Etablierung von *so* als Fokusmarker in der multiethnischen Population hinweisen.

¹⁷ Wir danken Manfred Krifka für diesen Hinweis.

¹⁸ Auf eine mögliche syntaktische Stützung für das Auftreten von Fokuspartikeln in zirkum-Position verweist Barbiere (ersch.), der annimmt, dass Fokuspartikeln grundsätzlich verdoppelt werden und sowohl an die Fokus-Phrase (DP, PP) als auch an TP treten. Bei Bewegung der Fokusphrase nach links wird die niedrigere Fokuspartikel mitbewegt und kann von der höheren absorbiert werden. Geschieht dies nicht, erhält man Konstruktionen mit zwei identischen Fokuspartikeln, etwa für das niederländische *maar* 'nur' in gesprochener Sprache. Die Analyse wird allerdings auf semantischer Ebene durch eine Charakterisierung von Fokuspartikeln als Quantifizierer motiviert, die auf einen genuinen, semantisch gebleichten Fokusmarker wie *so* nicht übertragbar wäre.

Im folgenden Abschnitt stellen wir nun ein Phänomen vor, das möglicherweise auf Kiezdeutsch beschränkt ist, d.h. wir haben hier bislang nicht nur quantitative, sondern kategoriale Unterschiede zwischen Haupt- und Ergänzungskorpus gefunden.

4.2 Wortstellungsvariation in der linken Satzperipherie

Die zweite hier exemplarisch vorgestellte Korpusstudie betrifft die Satzebene, nämlich die Wortstellungsvariation in der linken Satzperipherie deklarativer Hauptsätze, einer syntaktischen Domäne, die sich in der Forschung als sensitiv für die Auszeichnung informationsstruktureller Kategorien erwiesen hat (vgl. u.a. Rizzi 1997). Neben der kanonischen V2-Stellung (15a) finden sich im multiethnischen Hauptkorpus auch deklarative Hauptsätze mit V1-Stellung des Finitums (15b) sowie solche, in denen zwei Konstituenten dem finiten Verb vorangehen (15c).

- (15a) früher hat hxxx MEHR gelacht (KiDKo, Transkript MuH9WT)
 (15b) guckst du =n bisschen TRAurisch (KiDKo, Transkript MuH13MT)
 (15c) GESTern isch war KUdamm (KiDKo, Transkript MuH9WT)

Wir nehmen an, dass es in Kiezdeutsch aufgrund der Lockerung grammatischer Restriktionen zu freieren Besetzungsmöglichkeiten der linken Satzperipherie kommt, die eine direktere Abbildung informationsstruktureller Präferenzen erlauben (vgl. Wiese 2009, 2011a; Wiese/Freywald/Mayr 2009). Die hier vorgestellte Analyse der Korpusdaten konzentriert sich auf deklarative Hauptsätze vom Typ (15c), in denen zwei Konstituenten vor dem Finitum identifizierbar sind, und zwar Adverbial und Subjekt (Adv SVO).

Da für das Korpus zurzeit noch keine syntaktischen Annotationen vorliegen, basieren die Untersuchungen auf einer manuell gezogenen, unausgewogenen Stichprobe aus dem multiethnischen Hauptkorpus. Hierzu wurden aus den Aufnahmen von jedem/r Anker-Sprecher/in 30 deklarative Hauptsätze in den Transkripten manuell nach Vorfeldern mit zwei Konstituenten durchsucht. Wenn für einen Informanten/eine Informantin Belege in einem Ausschnitt gefunden wurden, wurden die kompletten Transkripte für diese/n Anker-Sprecher/in nach mehrfachen Vorfeldern vom Typ Adv SVO durchsucht. Als Adverbial zählten dabei Phrasen des Typs AdvP, PP, CP, die nicht als Argument des jeweiligen Verbs gewertet werden konnten. Drei Adverbiale traten mehrfach auf, nämlich *danach*, *nachher* und *gestern*. Vor diesem Hintergrund wurde dann über EXAKT gezielt nach mehrfachen Besetzungen der linken Satzperipherie gesucht, in denen satzinitial die Konstituente *danach*, *nachher* oder *gestern* steht. Dieses zweistufige Vorgehen ergab insgesamt 55 Belege für eine Mehrfachbesetzung der linken Satzperipherie im multiethnischen Hauptkorpus (davon: *danach* = 18 Belege, *nachher* = 5 Belege, *gestern* = 3 Belege); (16) zeigt zwei Beispiele.

- (16a) dann die sind zur Ubahn gerannt (KiDKo, Transkript MuH9WT)
 (16b) danach er er sacht so zu oxxx
 ey verPISS dich verPISS dich von mir (-).
 danach er sagt zu oxxx, geh mal WEG. (KiDKo, Transkript MuH9WT)

Um zu ermitteln, in welchem Verhältnis dieser Befund zu den monoethnischen Vergleichsdaten steht, wurde dort für die drei Adverbiale *danach*, *nachher* und *gestern* ebenfalls eine Gesamterhebung durchgeführt. Diese lieferte insgesamt 69 Vorkommen dieser Adverbiale im Korpus, davon jedoch keinen Beleg für Adv SVO.

Vergleicht man das Adv SVO-Muster mit Herausstellungskonstruktionen, wie z.B. Linksversetzung (17) und Hanging Topic (18), die auch in standardnaher gesprochener Sprache vorhanden sind, so sind die im multiethnischen Korpus gefundenen Belege für Adv SVO klar davon abzugrenzen. (17) und (18) fordern im Kernsatz ein resumptives Element (vgl. u.a. Altmann 1981; Frey 2004), das in den Adv SVO-Sätzen der Stichprobe nicht vorkommt.

(17) Diese Kälte, die ertrage ich nicht.

(18) Diese Kälte, ich ertrage die/sie nicht.

Die Korpus-Belege zu Adv SVO unterscheiden sich strukturell ebenfalls grundlegend von Sätzen wie in (19):

- (19) Licht in den Rollenklischee-Dschungel wollte die Bremer Biologin Beate Fox bringen.
 (taz Bremen, 15.7.2004; zit. in Müller 2005, S. 181)

Hier steht zwar ebenfalls mehr als ein Satzglied vor dem finiten Verb. Jedoch ist es konstitutiv für die Konstruktion in (19), dass das Subjekt typischerweise gerade nicht – wie bei Adv SVO – präverbal, sondern im Mittelfeld erscheint, während im Vorfeld mindestens ein anderes Argument des Verbs steht. Syntaktisch werden Vorfelder wie in (19) als Voranstellung einer (einzigen) verbalen Projektion mit leerem Kopf analysiert (Müller 2003, 2005), sie werden daher als sog. „scheinbare mehrfache Vorfeldbesetzung“ betrachtet.

Linearisierungen der Form Adv SVO sind auch aus dem Fremdspracherwerb bekannt (vgl. u.a. Clahsen 1984; Pienemann 1998; Dimroth 2002). Auch wenn sich gewisse Einflüsse des L2-Erwerbs in Kiezdeutsch niedergeschlagen haben mögen, ist die Adv SVO-Struktur jedoch nicht allein auf solche Einflüsse zurückzuführen. Die bisherigen Korpusbefunde deuten zum einen darauf hin, dass deklarative Hauptsätze in Kiezdeutsch im Allgemeinen V2-Stellung aufweisen:

- (20) Spk1: es gibt auch noch ERDbeer.
 MuP1MK: erdbeer WILL ich doch gar nich. ich will ma DIS hier probiern.
 (KiDKo, Transkript MuP1MK)

Zudem bleibt auch bei mehrfacher Vorfeldbesetzung die Satzklammer erhalten:

- (21) danach wir ham uns so: TOTgelacht.
 und sie hat nur geSCHRIEN, (KiDKo, Transkript MuH17MA)

Dies weist darauf hin, dass in Kiezdeutsch die kanonische Grundstruktur deutscher Sätze nicht grundsätzlich aufgegeben wird, etwa im Sinne einer Entwicklung hin zu genereller SVO-Abfolge, sondern dass dieses neue Muster mit einer funktionalen Differenzierung einhergeht und nicht unsystematisch, etwa durch ungenügende Beherrschung der deutschen Grammatik, zustande kommt.

Auf Basis der Stichprobe können erste, an späteren Stufen des Korpus (vgl. Kapitel 5) noch zu erweiternde Befunde zu charakteristischen grammatischen und informationsstrukturellen Merkmalen des Adv SVO-Musters formuliert werden. Das satzinitiale Adverbial trägt häufig eine temporale oder lokale Semantik und wird in einer Vielzahl der Fälle in Form einer Adverbphrase (16a), hier wiederholt als (22), oder einer Präpositionalphrase (23) realisiert.

- (22) JA die meinte so , JA wo is RXXX und so ,
isch meinte sie is grad vor ZEHN minuten geGANgen ,
dann die sind zur Ubahn gerannt (KiDKo, Transkript MuH9WT)
- (23) A:Lle freun sich , alLe , eGA:L ,
auf EINmal , isch dachte mir es wir eins zu EINS stehn
(KiDKo, Transkript MuH28MK)

Jedoch können auch Phrasen anderer Kategorien die Adv-Stelle einnehmen, wie z.B. die CP in (24):

- (24) Exxx: sie hat (-) an die TA:fel alle geholt , (-) war SO ein test also ,
sie hat (die) WÖRter gesagt und wir müssen an die Tafel schreiben , (-)
(typisch) (-)
MuH17MA: un= DANN ?
Exxx: weil ich FRECH war , (-) sie hat mich zur Tafel geholt .
(KiDKo, Transkript MuH17MA)

Bei der präverbalen Konstituente handelt es sich um das Subjekt des Satzes, das überwiegend pronominal erscheint (vgl. (22) bis (24)), selten auch nicht-pronominal. Ein Beispiel für eine Realisierung als volle DP ist (25):

- (25) ENDspiel . NACHspielzeit , ne . für die türken . (-)
HEISST doch nur dass die (-) äh: ANfangsmannschaften immer ERSten tauschen .
nachher die türken ham doch wieder (-) ZEIT .
alles NACHzuholen . also EINzuholen .
(KiDKo, Transkript MuP1MK)

Die Beispiele in (24) und (25) geben erste Hinweise auf eine größere Vielfalt hinsichtlich der syntaktischen und semantischen Merkmale der beiden präverbalen Konstituenten. Diese differenzierteren Besetzungsmöglichkeiten deuten an, dass es sich bei der im Hauptkorpus belegten mehrfachen Vorfeldbesetzung um einen produktiven Prozess handelt.

Informationsstrukturell geben die initialen Adverbiale den zumeist temporalen oder lokalen Rahmen vor, in dem die folgende Aussage interpretiert werden soll, d.h. sie tragen die Funktion von Framesettern. Diese rahmende Eingrenzung kann sich sowohl auf propositionaler Ebene (vgl. (24) und (25)) als auch im Sinne einer Strukturierung von Erzählsequenzen auf Diskursebene bewegen (vgl. (16a) und (16b)). Das präfinite Subjekt trägt den Status eines Aboutness-Topiks, das im Diskurs gegeben ist. Damit existieren zwei Elemente innerhalb einer Äußerung, die an den linken Rand streben, wobei Framesetter universell die absolute Initialposition präferieren (Jacobs 2001; Krifka 2008). Diese Präferenz kann in Kiezdeutsch offenbar eher realisiert werden als im Standarddeutschen. Die Abgrenzung der topikalischen Domäne vom Kommentarbereich durch das finite Verb ist mit Adv SVO direkter umgesetzt.

Unter Annahme dieser informationsstrukturellen Motiviertheit könnte es sich bei Adv SVO um den funktionalen Ausbau eines standarddeutschen strukturellen Modells handeln, nämlich die Voranstellung von Sprechaktsadverbialen (26) bzw. Sprechaktkonditionalen (27) vor den Gesamtsatz, die im Standarddeutschen ebenfalls ohne resumptives Element konstruiert wird (vgl. u.a. Günthner 1999; Meinunger 2009).

- (26) Ehrlich, ich bin von dir total enttäuscht. (Meinunger 2009, S. 116)
- (27) Wenn ich ehrlich bin (.) also (-) das reizt mich also gar net. (Günthner 1999, S. 217)

Meinunger (2009) geht unter Bezug auf Haegemans Ansatz zu „syntactic orphans“ (Haegeman 2009) davon aus, dass *ehrllich* hier nicht Teil der Satzbedeutung ist, sondern eine basisgenerierte Position oberhalb des Vorfelds in Form einer CP-Adjunktion besetzt. Dabei bezieht das Adverbial *ehrllich* seine diskursbezogene Bedeutung ausschließlich aus dieser Voranstellung (im Unterschied zur komplexeren Form *ehrllich gesagt*, die in gleicher Funktion auch im Mittelfeld erscheinen kann). Für die Adv SVO-Struktur in Kiezdeutsch lässt sich daraus die Hypothese ableiten, dass hier die im standardnahen Deutsch vorhandene strukturelle Möglichkeit der CP-Adjunktion genutzt und funktional weiter differenziert wird für eine informationsstrukturelle Auszeichnung als Framesetter bzw. für eine diskursstrukturelle Markierung durch das satzinitiale Element. Sowohl Framesetter als auch diskursbezogene Elemente haben Skopus über die gesamte Äußerung (die sie ja auf spezifische Weise im Diskurs bzw. in der Diskurswelt verankern); dies macht eine Analyse als Adjunktion an den höchsten Knoten äußerst plausibel.

Auch wenn Adv SVO insgesamt ein breiteres funktionales Spektrum abdeckt, weist es in seiner Funktion als diskursstrukturierende Markierung Parallelen zur Voranstellung von Diskursmarkern auf (28). Diese Parallelen könnten die Ausweitung des Adv SVO-Musters weiter stützen.

- (28) Also, ich gehe jetzt besser nach Hause.

Funktional dienen diese der Strukturierung des Diskurses und der Organisation der Sprecher-Adressaten-Interaktion. Sie sind semantisch gebleicht sowie syntaktisch und prosodisch nicht integriert (vgl. u.a. Schiffrin 1987; Auer/Günthner 2005). Im Unterschied zu *also* ist bei der diskursstrukturierenden Verwendung von *dann* und *danach* in (16) jedoch nicht von einer semantischen Bleichung auszugehen. Vielmehr machen *dann* und *danach* die temporale Abfolge der Erzählsequenzen explizit.

Die bisherigen Befunde weisen damit auf ein produktives neues Muster Adv SVO in Kiezdeutsch hin, das die im Deutschen angelegten Möglichkeiten ausbaut und für informations- und diskursstrukturelle Mittel weiter nutzbar macht. Dies wird nach weiterer Aufbereitung des Korpus anhand einer größeren syntaktisch aufbereiteten Datenbasis zu überprüfen sein.

5. Geplante weitere Aufbereitung

Um das Korpus für linguistische Abfragen auf mehreren Ebenen systematisch durchsuchbar zu machen, ist die automatisch gestützte weitere Aufbereitung der Daten vorgesehen, die mehrere Annotationsebenen liefern wird und es unter anderem ermöglicht, auf Wortarten, Phrasen und topologische Felder zu rekurren. Die aufbereiteten Korpusdaten werden in das Such- und Visualisierungswerkzeug ANNIS („ANNotation of Information Structure“) importiert (vgl. Zeldes et al. 2009). Im Unterschied zu einzelnen Annotationswerkzeugen, in denen nur die für das jeweilige Tool spezifischen Ebenen abbildbar sind, handelt es sich bei ANNIS um ein mächtiges korpuslinguistisches Werkzeug, das es ermöglicht, Annotationen verschiedener Tools und unterschiedlicher linguistischer Ebenen in einer gemeinsamen Darstellung zu visualisieren und so linguistische Phänomene auf unterschiedlichen Ebenen gleichzeitig zu betrachten. Außerdem ermöglicht die Anfragesprache AQL (ANNIS Query Language) eine einheitliche Durchsuchbarkeit unterschiedlicher Korpora.

Als Basis für die weitere Aufbereitung wird zunächst eine orthografische Normalisierung der Transkriptionen vorgenommen, auf der ein (semi-)automatischer PoS-Tagger operieren wird, der mit Hilfe eines vordefinierten Tagsets Wortarten („Parts of Speech“) identifiziert. Um für gesprochene Sprache charakteristische Merkmale (bspw. Klitika, gefüllte Pausen), die im Zuge der Normalisierung aufgelöst werden, später im Korpus wieder auffinden zu können, ist eine gesonderte Annotationsebene *Macro* vorgesehen, auf der die relevanten Spannen markiert werden.

Das PoS-Tagging bildet die elementare Grundlage für die syntaktischen Aufbereitungsschritte. Es werden zunächst die in den Äußerungen enthaltenen tiefsten Phrasen identifiziert, was Anfragen über die lexikalische Ebene hinaus an größere Chunks ermöglicht. Dies ist besonders für die Besetzung bestimmter syntaktischer Felder, wie bspw. der linken Satzperipherie, von zentraler Bedeutung. Um die topologischen Domänen der Äußerungen genauer eingrenzen und gezieltere Anfragen an diese Domänen formulieren zu können, wird auf das Chunking der tiefsten Phrasen dann das Parsen topologischer Felder aufbauen.

Als weitere Annotationsebenen sind die Markierung und Transkription fremdsprachlichen Materials sowie die Erstellung von damit alignierten Übersetzungen vorgesehen. Darüber wird es z.B. möglich sein, einen Einblick in das Auftreten von Code-Switching oder den Gebrauch von Mustern, die für das Deutschlandtürkische charakteristisch sind, zu bekommen.¹⁹

Belegstellen innerhalb der Transkripte sollen jederzeit auf die entsprechende Sequenz in den Audioaufnahmen bezogen werden können. An die Stelle der gegenwärtigen Alignierung (Abstände von ca. zwei Minuten) wird daher eine Feinalignierung von Audio- und Transkriptionsdateien treten.

Zum derzeitigen Bestand spontansprachlicher Daten (multiethnisches Hauptkorpus und monoethnisches Ergänzungskorpus) kommen künftig zwei Zusatzkorpora mit kontrolliert erhobenen, elizitierten Sprachdaten hinzu. Das Zusatzkorpora „Schriftliche Textproduktion“ wird schriftliche Produktionsdaten aus der multiethnischen und der monoethnischen Sprechergemeinschaft enthalten sowie darüber hinaus schriftliche Daten von Sprecher/innen aus der Türkei mit Türkisch als Muttersprache und Deutsch als gesteuert erworbener Fremdsprache (die Daten beinhalten die schriftliche Nacherzählung einer wortlosen Bildergeschichte, der sog. *Frog Story*, Mayer 1969; vgl. Freywald et al. 2010). Die Auswahl türkischer Deutschlerner/innen ist dadurch motiviert, dass Jugendliche türkischer Herkunft einen verhältnismäßig großen Anteil an der Gruppe der Kiezdeutschsprecher/innen stellen und einige Phänomene in Kiezdeutsch, wie bereits angedeutet, zumindest oberflächennahe Parallelen zu Lernervarietäten aufweisen (bspw. der Wortstellungstyp Adv SVO). Als zweites Zusatzkorpora werden mündliche Daten aus der multiethnischen und der monoethnischen Sprechergemeinschaft Berücksichtigung finden, die mit Hilfe des *Questionnaire on Information Structure* (Skopeteas et al. 2006) gezielt eliziert wurden; es handelt sich dabei um die Beschreibung einer Filmsequenz. Die beiden Zusatzkorpora ermöglichen damit (a) Studien zum sprachlichen Repertoire innerhalb einer Sprechergruppe (verschiedene konzeptuelle Äußerungsmodi, verschiedene soziale Register), (b) die Gegenüberstellung der sprachlichen Merkmale in elizitierten schriftlichen und

¹⁹ Zur Entstehung neuer Varietäten des Türkischen in Nordwesteuropa vgl. Boeschoten (2000), Rehbein (2001); speziell zum Deutschlandtürkischen vgl. Cindark/Aslan (2004) sowie Pfaff (1991).

mündlichen Äußerungen von multi- vs. monoethnischer Sprechergruppe und (c) den Vergleich von Kiezdeutsch-Daten mit DaF-Daten türkischsprachiger Deutschlerner/innen derselben Altersgruppe.

Um das KiezDeutsch-Korpus auch für andere Wissenschaftler/innen zugänglich zu machen, ist eine rechtlich abgesicherte wissenschaftliche Öffnung des Korpus vorgesehen, die auf weitergehenden Abstimmungen mit den (mittlerweile volljährigen) Sprecher/innen basiert. Hierfür wird ein zentraler lokaler Nutzerzugang eingerichtet, der durch eine öffentliche Online-Dokumentation aller Meta-Informationen zum Korpus begleitet wird.

Als Ergebnis wird das KiDKo eine wissenschaftlich allgemein zugängliche neue empirische Ressource für die Erforschung von Fragen zu Sprachstruktur, Sprachvariation, Sprachgebrauch und Sprachkontakt liefern: ein umfangreiches, automatisch durchsuchbares, multimodales digitales Korpus urbaner Jugendsprache im multiethnischen und monoethnischen Raum auf der Grundlage informeller Sprache in Peer-Group-Situationen.

Literatur

- Altmann, Hans (1981): Formen der „Herausstellung“ im Deutschen. Rechtsversetzung, Linksversetzung, Freies Thema und verwandte Konstruktionen. Tübingen: Niemeyer. (= Linguistische Arbeiten 106).
- Androutsopoulos, Jannis (1998): Deutsche Jugendsprache. Untersuchungen zu ihren Strukturen und Funktionen. Frankfurt am Main: Lang. (= VarioLingua 6).
- Androutsopoulos, Jannis (2007): Ethnolekte in der Mediengesellschaft Stilisierung und Sprachideologie in Performance, Fiktion und Metasprachdiskurs. In: Fandrych, Christian/Salverda, Reinier (Hg.): Standard, Variation und Sprachwandel in germanischen Sprachen/Standard, Variation and Language Change in Germanic Languages. Tübingen: Narr. S. 113-155. (= Studien zur deutschen Sprache 41).
- Appel, René (1999): Straattaal. De mengtaal van jongeren in Amsterdam. Toegepaste Taalwetenschap in Artikelen 62, S. 39-55.
- Auer, Peter (2003): „Türkenslang“: Ein jugendsprachlicher Ethnolekt des Deutschen und seine Transformationen. In: Häcki Buhofer, Annelies (Hg.): Spracherwerb und Lebensalter. Tübingen: Francke. S. 255-264. (= Basler Studien zur deutschen Sprache und Literatur 83).
- Auer, Peter (2006): Construction Grammar meets Conversation: Einige Überlegungen am Beispiel von „so“-Konstruktionen. In: Günthner, Susanne/Imo, Wolfgang (Hg.): Konstruktionen in der Interaktion. Berlin/New York: de Gruyter. S. 291-314. (= Linguistik 20).
- Auer, Peter (2007): Syntax als Prozess. In: Hausendorf, Heiko (Hg.): Gespräch als Prozess. Linguistische Aspekte der Zeitlichkeit verbaler Interaktion. Tübingen: Narr. S. 95-124. (= Studien zur deutschen Sprache 37).
- Auer, Peter (2010): Zum Segmentierungsproblem in der Gesprochenen Sprache. InLiSt – Interaction and Linguistic Structures 49. <http://www.inlist.uni-bayreuth.de/issues/49/index.htm> (24.1.2012).
- Auer, Peter/Dirim, İnci (1999): Das versteckte Prestige des Türkischen. Zur Verwendung des Türkischen in gemischtethnischen Jugendlichengruppen in Hamburg. In: Gogolin, Ingrid/Nauck, Bernhard (Hg.): Migration, gesellschaftliche Differenzierung und Bildung. Resultate des Forschungsschwerpunktprogramms FABER. Opladen: Leske & Budrich. S. 97-112.
- Auer, Peter/Günthner, Susanne (2005): Die Entstehung von Diskursmarkern im Deutschen – ein Fall von Grammatikalisierung? In: Leuschner, Torsten/Mortelmans, Tanja/De Groodt, Sarah (Hg.): Grammatikalisierung im Deutschen. Berlin/New York: de Gruyter. S. 335-362. (= Linguistik 9).
- Barbiers, Sjef (ersch.): Syntactic doubling and deletion as a source of variation. In: Picallo, Carme (Hg.): Linguistic Variation in the Minimalist Framework. New York: Oxford University Press.

- Bijvoet, Ellen (2003): Attitudes towards "Rinkeby Swedish", a group variety among adolescents in multilingual suburbs. In: Fraurud, Kari/Hyltenstam, Kenneth (Hg.): Multilingualism in Global and Local Perspectives. Papers from the 8th Nordic Conference on Bilingualism, November 1-3, 2001, Stockholm – Rinkeby, Stockholm: Centre for Research on Bilingualism & Rinkeby Institute of Multilingual Research. S. 307-316.
- Boeschoten, Hendrik (2000): Convergence and divergence in migrant Turkish. In: Mattheier, Klaus (Hg.): Dialect and Migration in a Changing Europe. Frankfurt am Main: Lang. S. 145-154. (= VarioLingua 12).
- Chambers, Jack K./Trudgill, Peter (1998): Dialectology. 2. Auflage. Cambridge: Cambridge University Press.
- Cheshire, Jenny/Fox, Sue/Kerswill, Paul/Torgersen, Eivind (2008): Ethnicity, friendship network and social practices as the motor of dialect change: linguistic innovation in London. In: Sociolinguistica 22, S. 1-23.
- Cheshire, Jenny/Fox, Sue/Kerswill, Paul/Torgersen, Eivind (2011): Contact, the feature pool and the speech community: The emergence of Multicultural London English. In: Journal of Sociolinguistics 15, S. 151-196.
- Cindark, Ibrahim/Aslan, Sema (2004): Deutschlandtürkisch? Ms. Mannheim: Institut für Deutsche Sprache. <http://www.ids-mannheim.de/prag/soziostilistik/Deutschlandtuerkisch.pdf> (24.1.2012).
- Clahsen, Harald (1984): The acquisition of German word order: a test case for cognitive approaches to L2 development. In: Andersen, Roger W. (Hg.): Second Languages. A Cross-linguistic Perspective. Rowley, MA: Newbury House. S. 219-242.
- Clyne, Michael (2000): Lingua franca and ethnolects in Europe and beyond. In: Sociolinguistica 14, S. 83-89.
- Cornips, Leonie (2004): Straattaal: Sociale betekenis en morfo-syntactische verschijnselen. In: De Caluwe, Johan et al. (Hg.): Taeldeman, man van de taal, schatbewaarder van de taal. Liber amicorum voor Johan Taeldeman. Gent: Academia. S. 175-188.
- Dirim, İnci/Auer, Peter (2004): Türkisch sprechen nicht nur die Türken. Über die Unschärfebeziehung zwischen Sprache und Ethnie in Deutschland. Berlin/New York: de Gruyter. (= Linguistik – Impulse & Tendenzen 4).
- Dimroth, Christine (2002): Topics, assertions, and additive words: How L2 learners get from information structure to target-language syntax. In: Linguistics 40, S. 891-923.
- Ehlich, Konrad (1987): *so* – Überlegungen zum Verhältnis sprachlicher Formen und sprachlichen Handelns, allgemein und an einem widerspenstigen Beispiel. In: Rosengren, Inger (Hg.): Sprache und Pragmatik. Lunder Symposium 1986. Stockholm: Almqvist & Wiksell. S. 279-298.
- Eksner, H. Julia (2006): Ghetto ideologies, Youth Identities and Stylized Turkish German. Turkish Youths in Berlin-Kreuzberg. Berlin: Lit Verlag. (= Spektrum 91).
- Fiehler, Reinhard (2003): Was sind die Grundeinheiten gesprochener Sprache? Ein altes Problem und ein neuer Lösungsvorschlag. In: Sprachtheorie und germanistische Linguistik 13, S. 145-72.
- Foster, Pauline/Tonkyn, Alan/Wigglesworth, Gillian (2000): Measuring spoken language: A unit for all reasons. In: Applied Linguistics 21, S. 354-375.
- Fraurud, Kari (2003): Svenskan i Rinkeby och andra flerspråkiga bostadsområden. In: Breivik, Torbjørn (Hg.): Språk i Norden 2003. Oslo: Novus. S. 62-89.
- Frey, Werner (2004): Notes on the syntax and the pragmatics of German Left Dislocation. In: Lohnstein, Horst/Trissler, Susanne (Hg.): The Syntax and Semantics of the Left Periphery. Berlin/New York: Mouton. S. 203-233. (= Interface Explorations 9).
- Freywald, Ulrike/Mayr, Katharina/Schalowski, Sören/Wiese, Heike (2010): Linguistic Fieldnotes II: Information Structure in Different Variants of Written German. Potsdam: Universitätsverlag Potsdam. (= Interdisciplinary Studies on Information Structure 14). <http://pub.ub.uni-potsdam.de/volltexte/2010/3683/> (24.2.2012).

- Freywald, Ulrike/Mayr, Katharina/Özçelik, Tiner/Wiese, Heike (2011): Kiezdeutsch as a multiethnolect. In: Kern, Friederike/Selting, Margret (Hg.): Ethnic Styles of Speaking in European Metropolitan Areas. Amsterdam/Philadelphia: Benjamins. S. 45-73. (= Studies in Language Variation 8).
- Füglein, Rosemarie (2000): Kanak Sprak. Eine ethnolinguistische Untersuchung eines Sprachphänomens im Deutschen. Universität Bamberg. Diplomarbeit.
- Ganuza, Natalia (2008): Syntactic variation in the Swedish of adolescents in multilingual urban settings. Subject-verb order in declaratives, questions and subordinate clauses. Stockholm: Centre for Research on Bilingualism, University of Stockholm. Dissertation.
- Golato, Andrea (2000): An innovative German quotative for reporting on embodied actions: *Und ich so / und er so* 'and I'm like/and he's like'. In: Journal of Pragmatics 32, S. 29-54.
- Günthner, Susanne (1999): *Wenn*-Sätze im Vor-Vorfeld: Ihre Formen und Funktionen in der gesprochenen Sprache. In: Deutsche Sprache 27, S. 209-235.
- Haegeman, Liliane (2009) [1991]: Parenthetical adverbials: the radical orphanage approach. In: Shaer, Benjamin et al. (Hg.): Dislocated Elements in Discourse: Syntactic, Semantic, and Pragmatic Perspectives. New York: Routledge. S. 331-347. (= Routledge Studies in Germanic Linguistics 12).
- Jacobs, Joachim (2001): The dimensions of topic-comment. In: Linguistics 39, S. 641-681.
- Kallmeyer, Werner/Keim, Inken (2002): Eigenschaften von sozialen Stilen der Kommunikation: Am Beispiel einer türkischen Migrantinnengruppe. In: Osnabrücker Beiträge zur Sprachtheorie 65, S. 35-56.
- Keim, Inken (2001): Die Powergirls – Aspekte des kommunikativen Stils einer Migrantinnengruppe aus Mannheim. In: Jakobs, Eva-Maria/Rothkegel, Annely (Hg.): Perspektiven auf Stil. Tübingen: Niemeyer. S. 375-400. (= Reihe Germanistische Linguistik 226).
- Keim, Inken (2007): Socio-cultural identity, communicative style, and their change over time: A case study of a group of German-Turkish girls in Mannheim/Germany. In: Auer, Peter (Hg.): Style and Social Identities. Alternative Approaches to Linguistic Heterogeneity. Berlin/New York: Mouton. S. 155-186. (= Language, Power and Social Process 18).
- Keim, Inken/Androutsopoulos, Jannis (2000): Hey Lan, isch geb dir konkret Handy. Deutsch-türkische Mischsprache und Deutsch mit ausländischem Akzent: Wie Sprechweisen der Straße durch die Medien populär werden. In: Frankfurter Allgemeine Zeitung 21, 26.1.2000.
- Kern, Friederike/Selting, Margret (2006): Einheitenkonstruktion im Türkendeutschen: Grammatische und prosodische Aspekte. In: Zeitschrift für Sprachwissenschaft 25, S. 239-272.
- Kirschbaum, Ilja (2002): *Schrecklich nett und voll verrückt*. Muster der Adjektiv-Intensivierung im Deutschen. Heinrich-Heine-Universität Düsseldorf. Dissertation.
- Kotsinas, Ulla-Britt (1992): Immigrant adolescents' Swedish in multicultural areas. In: Palmgren, Cecilia/Lövgren, Karin/Bolin, Goran (Hg.): Ethnicity in Youth Culture. Stockholm: University of Stockholm. S. 43-62.
- Kotsinas, Ulla-Britt (1998): Language contact in Rinkeby, an immigrant suburb. In: Androutsopoulos, Jannis/Scholz, Arno (Hg.): Jugendsprache – langue des jeunes – youth language. Linguistische und soziolinguistische Perspektiven. Frankfurt am Main: Lang. S. 125-148. (= VarioLingua 7).
- Krifka, Manfred (2008): Basic notions of information structure. In: Acta Linguistica Hungarica 55, S. 243-276.
- Mayer, Mercer (1969): Frog, where are you? New York: Dial.
- Meinunger, André (2009): Leftmost peripheral adverbs and adjectives in German. In: Journal for Comparative Germanic Linguistics 12, S. 115-135.
- Merkel, Silke/Schmidt, Thomas (2009): Korpora gesprochener Sprache im Netz – eine Umschau. In: Gesprächsforschung 10, S. 70-93. <http://www.gespraechsforschung-ozs.de/heft2009/px-merkel.pdf> (24.1.2012).

- Müller, Stefan (2003): Mehrfache Vorfelddbesetzung. In: Deutsche Sprache 31, S. 29-62.
- Müller, Stefan (2005): Zur Analyse der scheinbar mehrfachen Vorfelddbesetzung. In: Linguistische Berichte 203, S. 297-330.
- Nortier, Jacomine (2001): "Fawaka, what's up?" Language use among adolescents in Dutch mono-ethnic and ethnically mixed groups. In: Hvenekilde, Anne/Nortier, Jacomine (Hg.): Meetings at the Crossroads. Studies of Multilingualism and Multiculturalism in Oslo and Utrecht. Oslo: Novus. S. 61-73.
- Pfaff, Carol W. (1991): Turkish in contact with German: Language maintenance and loss among immigrant children in Berlin (West). In: International Journal of the Sociology of Language 90, S. 97-129.
- Pienemann, Manfred (1998): Language Processing and Second Language Development: Processability Theory. Amsterdam/Philadelphia: Benjamins. (= Studies in Bilingualism 15).
- Pittner, Karin/Berman, Judith (2006): *video ist echt schrott aber single ist hammer* – Jugendsprachliche Nomen-Adjektiv-Konversion in der Prädikativposition. In: Deutsche Sprache 34, S. 233-250.
- Quist, Pia (2000): Ny københavnsk "multiethnolect". Om sprogbrug blandt unge i sprogligt og kulturelt heterogene miljøer. In: Danske Talesprog 1, S. 143-211.
- Quist, Pia (2005): New speech varieties among immigrant youth in Copenhagen – a case study. In: Hinnenkamp, Volker/Meng, Katharina (Hg.): Sprachgrenzen überspringen. Sprachliche Hybridität und polykulturelles Selbstverständnis. Tübingen: Narr. S. 145-161. (= Studien zur deutschen Sprache 32).
- Quist, Pia (2008): Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. In: International Journal of Bilingualism 12, S. 43-61.
- Rehbein, Jochen (2001): Turkish in European societies. In: Lingua e Stile 36, S. 317-334.
- Rizzi, Luigi (1997): The fine structure of the left periphery. In: Haegeman, Liliane (Hg.): Elements of Grammar. Handbook in Generative Syntax. Dordrecht: Kluwer. S. 281-337.
- Schiffrin, Deborah (1987): Discourse Markers. Cambridge: Cambridge University Press (= Studies in Interactional Sociolinguistics 5).
- Schlobinski, Peter/Kohl, Gabi/Ludewigt, Irmgard (1998): Jugendspezifische Sprechweisen. München: Lincom Europa. (= Text- und Tonbibliothek sprachlicher Varietäten 2). [CD-ROM].
- Schmidt, Thomas/Wörner, Kai (2005): Erstellen und Analysieren von Gesprächskorpora mit EXMARaL-DA. In: Gesprächsforschung 6, S. 171-195. <http://www.gespraechsforschung-ozs.de/heft2005/px-woerner.pdf> (24.1.2012).
- Selting, Margret et al. (1998): Gesprächsanalytisches Transkriptionssystem (GAT). In: Linguistische Berichte 173, S. 91-122.
- Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung 10, S. 353-402. <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf> (24.1.2012).
- Skopeteas, Stavros et al. (2006): Questionnaire on Information Structure (QUIS): Reference Manual. Potsdam: Universitätsverlag Potsdam. (Interdisciplinary Studies on Information Structure 4). <http://opus.kobv.de/ubp/volltexte/2007/1241/> (24.1.2012).
- Svendsen, Bente A. (2010): Linguistic practices in multilingual urban contexts in Norway: An overview. In: Quist, Pia/Svendsen, Bente A. (Hg.): Multilingual Urban Scandinavia. New Linguistic Practices. Bristol: Multilingual Matters. S. 12-16.
- Tertilt, Hermann (1996): Turkish Power Boys. Ethnographie einer Jugendbande. Frankfurt am Main: Suhrkamp.
- Thurmair, Maria (2001): Vergleiche und Vergleichen. Eine Studie zu Form und Funktion der Vergleichsstrukturen im Deutschen. Tübingen: Niemeyer. (= Linguistische Arbeiten 433).
- Weinrich, Harald (2003): Textgrammatik der deutschen Sprache. Hildesheim: Olms.

- Wiese, Heike (2006): „Ich mach dich Messer“: Grammatische Produktivität in Kiez-Sprache. In: Linguistische Berichte 207, S. 245-273.
- Wiese, Heike (2009): Grammatical innovation in multiethnic urban Europe: New linguistic practices among adolescents. In: Lingua 119, S. 782-806.
- Wiese, Heike (2011a): The role of information structure in linguistic variation: Evidence from a German multiethnolect. In: Gregersen, Frans/Parrott, Jeffrey/Quist, Pia (Hg.): Language Variation – European Perspectives III. Amsterdam/Philadelphia: Benjamins. S. 83-96. (= Studies in Language Variation 8).
- Wiese, Heike (2011b): *so* as a focus marker in German. In: Linguistics 49, S. 991-1039.
- Wiese, Heike (2012): Kiezdeutsch. Ein neuer Dialekt entsteht. München: Beck.
- Wiese, Heike (ersch.): Das Potenzial multiethnischer Sprechergemeinschaften. In: Das Deutsch der Migranten. Berlin/New York: de Gruyter. (= Jahrbuch des Instituts für Deutsche Sprache 2012).
- Wiese, Heike/Duda, Sibylle (ersch.): A new German particle 'gib(t)s' – The dynamics of a successful cooperation. In: Spalek, Katharina/Domke, Juliane (Hg.): Sprachliche Variationen, Varietäten und Kontexte. Beiträge zu psycholinguistischen Schnittstellen. Tübingen: Stauffenburg.
- Wiese, Heike/Freywald, Ulrike/Mayr, Katharina (2009): Kiezdeutsch as a Test Case for the Interaction between Grammar and Information Structure. Potsdam: Universitätsverlag Potsdam. (= Interdisciplinary Studies on Information Structure 12). <http://opus.kobv.de/ubp/volltexte/2009/3837/> (24.1.2012).
- Zeldes, Amir/Ritz, Julia/Lüdeling, Anke/Chiaros, Christian (2009): ANNIS: A search tool for multi-layer annotated corpora. In: Proceedings of Corpus Linguistics 2009, July 20-23, Liverpool, UK.

Heike Wiese, Ulrike Freywald, Sören Schalowski, Katharina Mayr
 Universität Potsdam
 Institut für Germanistik
 Am Neuen Palais 10
 D-14469 Potsdam
 E-Mail: heike.wiese@uni-potsdam.de,
 E-Mail: ulrike.freywald@uni-potsdam.de,
 E-Mail: soeren.schalowski@uni-potsdam.de,
 E-Mail: katharina.mayr@uni-potsdam.de