

Die Datenbank Gesprochenes Deutsch (DGD) - Sammlung, Dokumentation, Archivierung und Untersuchung gesprochener Sprache als Aufgaben der Sprachwissenschaft

Reinhard Fiehler & Peter Wagener

1. Sammlung, Dokumentation und Archivierung gesprochener Sprache

1.1. Unterschiedliche Verfügbarkeit von geschriebener und gesprochener Sprache

"Ich saz uf eime steine und dahte bein mit beine..." Sie haben vergessen, wie Walther von der Vogelweides Gedicht "Der Wahlstreit" weitergeht? Kein Problem. Wenn nicht in ihrem eigenen Bücherschrank, finden sie die verschriftlichte Version seiner Lieder und Gedichte in Hunderten von Bibliotheken und natürlich auch im Internet.

Sie wollen Luthers Tischreden einmal hören oder in Erfahrung bringen, wie der Verkauf einer Kuh auf einem mittelalterlichen Markt abgewickelt wird? Fehlanzeige!

Anders als schriftliche Texte, die auf Dauerhaftigkeit angelegt sind, ist die gesprochene Sprache ein flüchtiges Medium: Einen Moment, nachdem etwas geäußert wurde, ist es verfliegen. Es bleiben nur die Spuren in den Gedächtnissen der Personen, die diese Äußerung gehört haben, und diese Spuren sind weder besonders genau und verlässlich, noch sind sie vergleichbar dauerhaft. Die Flüchtigkeit gesprochener Sprache schränkt die Möglichkeiten ihrer Sammlung, Archivierung und auch ihrer wissenschaftlichen Untersuchung in massiver Weise ein.

Erst mit der Entwicklung von Geräten zur akustischen und audiovisuellen Aufzeichnung veränderte sich die Zugänglichkeit von gesprochener Sprache auf eine qualitative Art und Weise. Die Entwicklung und Verbreitung von technischen Geräten zur Konservierung und Reproduktion von Gesprächen setzt mit Thomas Edisons Erfindung des Phonographen im Jahre 1877 und Emil Berliners Gramophon zehn Jahre später ein und setzt sich im 20. Jahrhundert mit Tonbandgeräten, Kassettenrekordern und Videokameras fort. Die technische Perfektionierung der Tontechnik und vor allem ihre fortschreitende Miniaturisierung und Digitalisierung in den letzten Jahrzehnten ist so eine wesentliche Voraussetzung für die massenhafte Sammlung und Archivierung von gesprochener Sprache und eine detaillierte wissenschaftliche Untersuchung von mündlicher Kommunikation. Die Möglichkeit zu solchen Untersuchungen besteht damit erst seit gut hundert Jahren. Setzt man eine bestimmte Ausgereiftheit und Verbreitung der Geräte voraus, kann man sagen, dass sie seit den 50er Jahren des 20. Jahrhunderts gegeben ist. Alles, was zuvor gesprochen wurde, und alles, was trotz der Verfügbarkeit entsprechender Geräte nicht aufgezeichnet wird, ist für eine genaue wissenschaftliche Untersuchung unwiederbringlich verloren. Die Erforschung gesprochener Sprache kann sich so im Wesentlichen nur auf die Gegenwartssprache und auf die Sprache der jüngsten Vergangenheit stützen.

Eine zweite wesentliche Voraussetzung ist die Entwicklung von Verfahren zur Verschriftlichung (Transkription) konservierter Gespräche. Transkriptionen ermöglichen eine Vergegenwärtigung und 'Betrachtung' der Äußerungen und Gespräche, wie sie allein durch das Abhören der Aufzeichnung nicht zu erreichen ist. Die Entwicklung der Transkriptionssysteme erfolgte Hand in Hand mit dem Einsatz der genannten Geräte für sprachwissenschaftliche Zwecke. Während in der ersten Hälfte des 20. Jahrhunderts Transkripte auch "kurrent", d.h. unmittelbar im Gespräch mit der Gewährsperson, angefertigt wurden – oft auch als Ersatz für eine Tonaufnahme wegen ihrer hohen Kosten –, wurden die Transkriptionskonventionen in den vergangenen Jahrzehnten immer weiter verfeinert und Transkriptionen sind heute ohne die wieder und wieder abhörbaren Tonaufnahmen gar nicht denkbar. Je nach Erkenntnisinteresse und dem Fokus der Analyse werden z.B. phonetische, prosodische oder gesprächsstrukturelle Merkmale der Interaktion in den Transkripten notiert. Die gesprochene Sprache wird mit Hilfe von Transkriptionen in einem bisher nicht gekannten Detaillierungsgrad aufbereitet und für die linguistische Untersuchung zugänglich.

So entstanden im Zuge der Entwicklung der Gesprochene-Sprache- und der Gesprächsforschung eine Vielzahl von Korpora, allerdings auf einem sehr unterschiedlichen Stand der Dokumentation und Archivierung. Auch ihre Zugänglichkeit variiert stark, da es sich zumeist um "Privatbestände" handelt.¹

Die Sammlung, Dokumentation, Archivierung und Tradierung mündlicher Kommunikation und gesprochener Sprache hat trotz der positiven Entwicklungen der letzten Jahrzehnte aber bei weitem noch nicht den Stand und die Standards erreicht, die für schriftliche Texte z.B. in Archiven oder Bibliotheken üblich sind.²

1.2. Archive gesprochener Sprache

Die Bedeutung der tontechnischen Erfindungen des 19. Jahrhunderts wurde schnell erfasst, die Möglichkeit, auf diese Weise das kollektive kulturelle Gedächtnis der Menschheit durch eine neue Dimension anzureichern, führte zu einer Gründungswelle von Phonogrammarchiven: Wien 1899, Paris 1900, London 1901, St. Petersburg 1903, Zürich 1909. Die Einrichtung des ersten Archivs, des "Phonogrammarchivs der Kaiserlichen Akademie der Wissenschaften in Wien", wurde 1899 beantragt und nach einer Erprobungsphase mit drei Sprachaufnahmenreisen (1901) Anfang des Jahres 1903 von der Akademie endgültig beschlossen. Den "Mitteilungen der Phonogrammarchivkommission Nr. 1" ist zu entnehmen, was als vordringlichste Aufgabe des Archivs gesehen wurde (zitiert nach Hajek (1937:222)):

Es wären zunächst die sämtlichen europäischen Sprachen in ihrem Zustande am Ende des 19. Jahrhunderts aufzunehmen. Anschließend wären die europäischen Dialekte und sodann im Verlaufe weiterer Jahrzehnte die sämtlichen Sprachen der Erde aufzunehmen.

Monumentale Ziele ganz im Geiste des positivistischen Zeitalters werden hier formuliert, gegen die sich die heutige Forschungspolitik eher kleinmütig ausnimmt. Denn viele der Großprojekte dieser Zeit wurden ja tatsächlich mit immensem Ar-

¹ Vgl. Wagener / Bausch (1997) und Glas / Ehlich (2000).

² Vgl. Schmidt (2005).

beitseifer realisiert. Man beachte die z.T. noch heute nicht überholten Ergebnisse von Lexikographie, Sprachkartographie, Textedition – und eben auch die Archivalien der Tonarchive.

Im deutschsprachigen Raum spielt neben Wien seit den 1920er Jahren vor allem Berlin eine wichtige Rolle. 1920 wurde die "Lautabteilung der Preußischen Staatsbibliothek" gegründet, fortgeführt ab 1933 als "Institut für Lautforschung" und der Universität Berlin angegliedert, zunächst mit ähnlichem Anspruch wie das Wiener Phonogrammarchiv. Komplementär dazu auf die deutsche Sprache beschränkt entstand in Berlin 1932 auch das "Deutsche Spracharchiv", das nach Zwischenstationen in Braunschweig, Münster und Bonn heute Teil des Instituts für Deutsche Sprache in Mannheim ist. Das Deutsche Spracharchiv stand auch Pate bei der Gründung zahlreicher kleiner Tonarchive, die seit den 1950er Jahren an den Universitäten entstanden – parallel zur Verfügbarkeit der Tontechnik, zunächst in Gestalt der "Heimtonbandgeräte". Schon bei den großen Aufnahmeaktionen des Deutschen Spracharchivs in den 50er und 60er Jahren zur Sammlung von deutschen Dialekten und Umgangssprachen waren viele einschlägige Abteilungen an den Universitäten beteiligt, die dann auch eigene Bestände anlegten. Dies und die Materialien aus zahlreichen empirischen Forschungsprojekten der letzten Jahrzehnte haben dazu geführt, dass es heute durchaus eine große Zahl von Tonaufnahmen mit gesprochenem Deutsch gibt, die aber weitgehend unerschlossen und z.T. unter unzureichenden Bedingungen an vielen verschiedenen Orten lagern.³

Eine Sonderstellung nimmt hier schon seit seiner Gründung das Deutsche Spracharchiv ein, das im Hinblick auf Aufnahmetechnik, Dokumentation und Archivierung stets hohe Standards einhalten konnte. Seit der Übergang zur digitalen Tontechnik und die Digitalisierung der Archivbestände vollzogen wird und vor allem seit die systematische computergestützte Erfassung und Erschließung der Bestände durch die Einrichtung der "Datenbank Gesprochenes Deutsch" begonnen und ein Stück weit entwickelt werden konnte, kann das Institut für Deutsche Sprache mit Fug und Recht auch als zentrale Dokumentations- und Archivierungsstelle für gesprochenes Deutsch angesehen werden.

Dennoch ist die Sammlung, Dokumentation und Archivierung gesprochener Sprache als kulturgeschichtliche Aufgabe und die sprachwissenschaftliche Untersuchung der gesprochenen Varianten des Deutschen und ihrer Verwendungsweisen in mancher Hinsicht nur Stückwerk. In der Vergangenheit wurde die Arbeit vor allem durch die Auswirkungen der beiden Weltkriege zerstört oder zurückgeworfen, in der Gegenwart wird der Stellenwert der Gesprochene-Sprache- und Gesprächsforschung – auch auf Grund der aufwändigen Bearbeitung des Materials und der damit verbundenen Kosten – oft nicht ausreichend gewürdigt. Aber eine Trendwende scheint sich anzubahnen, bei der dem IDS eine wichtige Funktion zukommen könnte. So kommt beispielsweise der Senatsausschuss Evaluierung der Leibniz-Gemeinschaft in seinem Bewertungsbericht vom 23.12.2003 zu folgender Empfehlung:

Der Aufbau von Text- und Sprachkorpora für das Deutsche ist eine wichtige Aufgabe von gesamtgesellschaftlichem Interesse, deren Bedeutung als Grundlage für die sprachwissenschaftliche Forschung und die Sprachtechnologie künftig noch

³ Vgl. Wagener/Bausch (1997).

zunehmen wird. Das IDS sollte diese Aufgabe noch deutlicher als bisher zentral wahrnehmen und die Korpusakquisition mit dem Ziel der Korpuserweiterung unbedingt forcieren. Ein dringendes Desiderat ist die Verfügbarkeit eines ausgewogenen nationalen Referenzkorpus, wie es das British National Corpus für das Englische darstellt. Langfristiges Ziel des IDS sollte es sein, auf ein solches Korpus hinzuwirken. [...] Wichtig für die externe Nutzung ist es zudem, dass das IDS Zugang und Nutzungsmöglichkeiten für externe Wissenschaftler/-innen erleichtert. (Bewertungsbericht Institut für Deutsche Sprache, S. 2)

2. Datenbank Gesprochenes Deutsch

Die Datenbank Gesprochenes Deutsch (DGD) ist die digitale Realisierungsform des Deutschen Spracharchivs (DSAv).⁴ Ihre Aufgabe ist, im Laufe der Zeit den Gesamtbestand des DSAv und weitere einzuwerbende Korpora digital zu archivieren und diese Materialien des Archivs in elektronischer Form für interne wie für externe Nutzer zu erschließen und variabel zugänglich zu machen. Die DGD erfüllt damit einerseits Archivfunktion,⁵ auf der anderen Seite ist sie ein Forschungsinstrument.

2.1. Bestandteile und aktueller Bestand

Das Deutsche Spracharchiv enthält im Wesentlichen drei Arten von Materialien: Das sind primär ca. 17 000 Tonaufnahmen, zu denen fast immer auch Metadaten erhoben worden sind, in der Regel also dokumentarische Informationen über den oder die Sprecher, den Gesprächstyp, die Aufnahmesituation u.ä. Organisiert sind die Tonaufnahmen in mehr als 30 Korpora,⁶ die aus verschiedenen Aufnahmeaktionen oder Forschungsprojekten stammen. Außerdem existieren zu vielen Tonaufnahmen Transkriptionen. Der Bestand setzt sich aus Tonaufnahmen mit deutschen Dialekten, Umgangssprachen, der Standardsprache und aus Aufnahmen zu den verschiedenen (Gesprächs-)Formen der verbalen Interaktion zusammen.

Innerhalb des IDS dient die DGD als Archiv, das die Tonaufnahmen, die Transkripte und die Dokumentationen auf neue Weise verwaltet und für die linguistische Arbeit z.B. in den Forschungsprojekten des IDS verfügbar macht. An PC-Workstations kann man Tonaufnahmen anhören, Transkripte lesen und die gesamte Funktionalität der Datenbank-Werkzeuge nutzen. Für die Nutzer außerhalb des IDS werden ausgewählte Materialien freigegeben. Der Zugang von außen erfolgt über eine Schnittstelle zum Internet, die Adresse lautet:

<http://dsav-oeff.ids-mannheim.de>

⁴ Die DGD wurde von Wolfgang Schneider (Dortmund) programmiert.

⁵ Dies ist keine triviale Aussage, denn ein virtuelles Archiv auf digitaler Basis, das Bestand haben will, bedarf eines systematischen Aufbaus. Die DGD basiert auf einem komplexen System stabiler und zukunftssicherer Archivformate. Damit sind auch unterschiedliche effiziente Nutzungsmöglichkeiten bereits angelegt. - Einige dieser Aspekte sind jetzt auch bei Schmidt (2005) thematisiert.

⁶ Eine Liste der Korpora findet sich in der DGD wie folgt: Rufen Sie die Startseite unter der aufgeführten URL auf und folgen Sie dem Link "Korpora" (Links sind im Folgenden so notiert: >Korpora). Von der auf diese Weise erreichten Seite aus gelangt man auch zu den Beschreibungen der einzelnen Korpora.

Über diese Adresse sind zwei Versionen der DGD zugänglich: Die Öffentliche Version und die Wissenschaftlerversion. Beide Versionen bieten die Möglichkeit, die Werkzeuge der Datenbank zu nutzen, also im Wesentlichen die Volltext-Recherche am Bildschirm, die an die Stelle der Suche in Papiermedien tritt. Eine Recherche in den dokumentarischen Daten gibt einen Überblick über die Bestände oder kann je nach der vom Nutzer formulierten Anforderung bestimmte Gruppen von Sprechern, bestimmte Gesprächstypen, Sprache aus verschiedenen Sprachlandschaften u.ä. suchen. Wenn Transkriptionen zu den Interaktionen vorhanden sind, können diese angezeigt werden. Die Transkripte wurden mit den zugehörigen digitalen Tonaufzeichnungen auf der Wortebene aligniert,⁷ so dass ein Navigieren in den Aufnahmen und eine effiziente Recherche möglich sind. Der Suchvorgang führt so bis zu den einzelnen Belegstellen für die untersuchten Phänomene. Die mit einer Tonaufnahme alignierten Transkripte kann man am Bildschirm nicht nur lesen, sondern auch (ausschnittweise) anhören. Die Tonwiedergabe wird dabei durch Mausklick auf eine beliebige Stelle im Transkript gestartet und liefert benutzerdefiniert einen Ausschnitt von jeweils 5 bis 15 Sekunden vor und nach der angeklickten Position. Diese Beschränkung auf maximal 30 Sekunden ist im Wesentlichen den derzeitigen Übertragungskapazitäten des Internets geschuldet. Die im Archiv als RIFF-WAVE-Dateien gespeicherten Aufnahmen werden aus diesem Grund für die ausschnittweise Präsentation über das Internet in eine datenreduzierte Form (wahlweise als mp3- oder wma-Datei) gewandelt.

Öffentliche und Wissenschaftlerversion unterscheiden sich in der Menge des zur Verfügung stehenden Materials. Die Öffentliche Version enthält zurzeit dokumentarische Daten zu ca. 9400 Aufnahmen und exemplarisch 10 alignierte Transkripte aus verschiedenen Korpora.

Die Wissenschaftlerversion der DGD, für die man sich von der Öffentlichen Version aus anmelden kann (Startseite >Datenbankrecherche >Zugang zur Datenbankrecherche >Zugangsberechtigung für die Wissenschaftlerversion beantragen), enthält zur Zeit dokumentarische Daten zu ca. 9400 Aufnahmen und Transkripte zu mehr als 3100 Aufnahmen, darunter (Stand: 07/2005)

- 207 alignierte Transkripte aus dem Freiburger Korpus (FR)
- 26 alignierte Transkripte aus dem Korpus Dialogstrukturen (DS)
- 398 alignierte Transkripte aus dem Pfeffer-Korpus (PF)
- 941 alignierte Transkripte aus dem Zwirner-Korpus (ZW)
- 9 alignierte Transkripte aus dem Korpus Deutsch in Israel (IS)

Die "Bestandsübersicht" (Startseite >Korpora >Bestandsübersicht) liefert weitere Details. Die alignierten Tonaufnahmen/Transkripte haben eine Gesamtdauer von mehr als 300 Std. und enthalten über zwei Mill. Wortformen. Die jetzt im Internet zugänglichen digitalen Transkripte enthalten insgesamt ca. sieben Mill. Wortformen. Die aktuelle Bestandsstatistik findet man unter Startseite >Korpora >Statistik.

⁷ Das Alignment ist ein im IDS weiter entwickeltes und für die DGD modifiziertes Verfahren zur weitgehend automatischen Synchronisierung von Tonaufnahmen und Transkripten, siehe: <http://www.ids-mannheim.de/ksgd/kt/alignment.html>

2.2. Nutzungsmöglichkeiten und Beispiele für Recherchen

Die Nutzer der Wissenschaftlerversion können die freigegebenen Dokumentationen, Transkripte und Tonausschnitte herunterladen.⁸ Diese Downloads sind kostenlos. Die Dokumentationen und Transkripte werden wahlweise als unformatierte oder formatierte Textdateien angeboten. Die Transkripte zusätzlich auch als Textgriddateien, was ihre Nutzung mit Hilfe des Analyseprogramms Praat ermöglicht.⁹

Auf diese Weise verändert sich der Service des Deutschen Spracharchivs grundlegend. Der Internet-Nutzer kann die bislang vom DSAV-Personal gelieferten Dienstleistungen direkt selbst auslösen. Das Personal wird damit frei für spezielle weitergehende Anforderungen. Nur wenn Kopien von kompletten Tonaufnahmen gewünscht werden, wird (vorläufig) weiterhin auf dem Postweg geliefert, und dieser Service ist auch weiterhin kostenpflichtig (Selbstkosten).

Lässt man sich über die erwähnte Internetadresse die Startseite der DGD auf dem Bildschirm ausgeben (<http://dsav-oeff.ids-mannheim.de>), erhält man einen ersten Überblick über die Struktur der DGD sowie eine Liste von Links, die zu Informationsseiten über den Aufbau der Datenbank, die Bestände, die Möglichkeiten der Recherche, die Geschichte des Deutschen Spracharchivs, den Service und die Erreichbarkeit des Personals führen. Das zentrale, auch farblich hervorgehobene Element der Startseite ist der Link, der zur Volltextrecherche in der Datenbank führt. Mit Hilfe einer Suchmaske (Startseite >Datenbankrecherche >Zugang zur Datenbankrecherche >Zugang zur...), in der eine ganze Reihe von Optionen gewählt werden können, werden nach den Vorgaben der Nutzer unterschiedlich gerichtete, umfassende oder auch eingeschränkte Suchvorgänge gestartet. Einige Beispiele wollen wir im Folgenden kurz beschreiben.

Das technische Werkzeug, das für die Suche eingesetzt wird, ist das im IDS entwickelte (und dort zunächst für die schriftlichen Korpora eingesetzte) Volltextretrievalinstrument COSMAS II, für das in der Datenbank Gesprochenes Deutsch eine Internet-Schnittstelle programmiert wurde.¹⁰ Die Volltextrecherche kann der Internetnutzer für alle Materialstränge der DGD einsetzen: für die Suche in den Metadaten, also in den Dokumentationen der einzelnen archivierten Interaktionen, für die Suche in den Sprachdaten, die in Form der Transkripte präsentiert werden, und über die Transkripte auch in den Tonaufnahmen, soweit sie mit den Transkripten bereits aligniert sind.

⁸ Nach einer Anmeldung zur erweiterten Nutzung – auf Formularen, die dem Nutzer der Wissenschaftlerversion per Email zugeschickt werden.

⁹ Zu Praat siehe: <http://www.praat.org>

¹⁰ Zu COSMAS II siehe: <http://www.ids-mannheim.de/cosmas2>

Beispiel 1: Suche in den Dokumentationen

Eine Suche in den Dokumentationen¹¹ kann z.B. so aussehen: Sie suchen Dialekt-aufnahmen mit westfälischem Niederdeutsch und wollen sich einen Überblick verschaffen, wie viele Aufnahmen das DSAv aus diesem Raum zu bieten hat. In der Suchmaske wählen Sie aus den Korpora die schon vorgegebene Kategorie "nur 'Dialekt-Korpora' auswählen" und im Feld "Mundarten" die Kategorie "Westniederdeutsch, Westfälisch". Die Recherche liefert Ihnen nach wenigen Sekunden eine Liste von 244 Fundstellen (siehe Abb.1; Stand: 07/2005); aufgeführt sind die Kennungen der Interaktionen mit einer Kurzbeschreibung über die Inhalte der Interaktion, das Jahr der Aufnahme und den Orten, die für den Sprecher eine Rolle gespielt haben (Geburtsort, Wohnorte u.ä.). Die Kennungen sind als Link ausgelegt und führen zu den detaillierteren Dokumentationen der Aufnahme. Von der Liste der Fundstellen oder auch von den Dokumentationen aus können Sie dann weiter zu den Transkripten gelangen, sofern sie für die jeweilige Interaktion bereits vorliegen bzw. zu den alignierten Transkripten, den Materialien, die Sie herunterladen können usw.

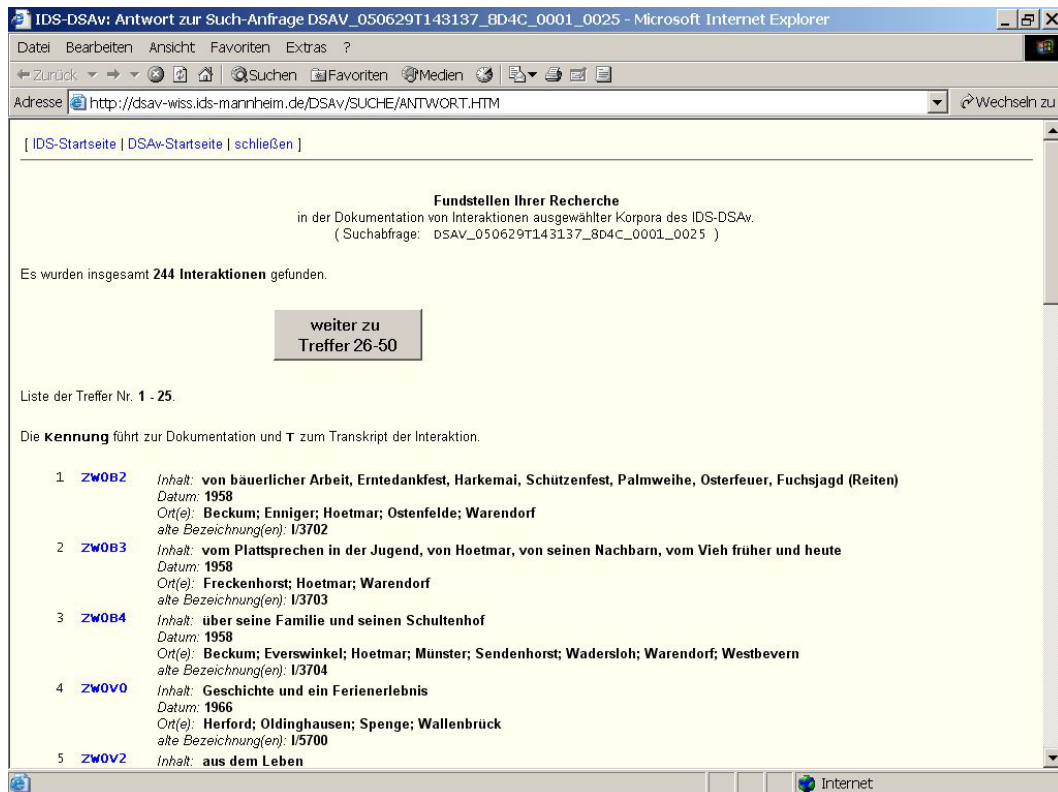


Abb. 1: Ausschnitt aus der Liste der Fundstellen zu Beispiel 1

¹¹ Siehe auch die Demo-Suchanfrage 1 am Ende der Suchmaske.

Beispiel 2

Wenn Sie an alten Mühlen oder dem Müllerhandwerk vor dem 2. Weltkrieg interessiert sind, könnten Ihre Eingaben in der Suchmaske so aussehen: Sie klicken das Kästchen "alle Korpora" an und tragen in dem Eingabefeld zur Suche in den Dokumentationen z.B. "Alter=50-99 + Geschlecht=männlich + Beruf=Müller" ein. Auf diese sehr spezielle Anfrage präsentiert die Suchmaschine 21 Interaktionen (Stand: 07/2005) mit Müllern bzw. ehemaligen Müllern, die (meistens in ihrem Dialekt) über ihre Mühlen oder ihre (frühere) Tätigkeit in Mühlen berichten.

Bei diesen Recherchen sind die Ergebnisse für Öffentliche und Wissenschaftlerversion übrigens identisch, weil beide, wie oben erwähnt, auf die Dokumentationen von ca. 9400 Aufnahmen zugreifen. Wollen Sie in konkretem Sprachmaterial recherchieren, sollten Sie sich der kurzen Anmeldeprozedur für die Wissenschaftlerversion unterwerfen, denn dort stehen Ihnen die ca. 3100 derzeit aufbereiteten Transkripte zur Verfügung, während die Öffentliche Version hier nur 10 Beispiele liefert.

Beispiel 3: Suche in den Transkriptionen

Eine Suche in den Transkriptionen¹² kann z.B. so aussehen: Sie möchten prüfen, welche Belege die DGD für aktuelle sprachwissenschaftliche Fragen liefert, z.B. für das Phänomen Verbzweitstellung in mit *weil* oder *obwohl* eingeleiteten Nebensätzen oder die landschaftlich differenzierte Verwendung von *eben*, *halt* oder der Kombination aus beiden Partikeln. Für die letztere Frage sei hier die Recherche vorgeführt.

Sie klicken in der Suchmaske "alle Korpora" an (denn die Auswahl von Korpora muss bei jeder Recherche erfolgen, damit die Suchmaschine weiß, wo sie suchen soll) und im Feld "(Suche) in Transkripten mittels Volltext-Recherche" das Suchwort "eben". Ergebnis: 8486 Fundstellen in Transkripten (Stand: 07/2005), von denen die meisten tatsächlich die Partikel *eben* präsentieren und nur einige wenige das Adjektiv. Die Belege werden in ihrem Textzusammenhang präsentiert, wobei standardmäßig insgesamt 100 Zeichen ausgegeben werden. Auf diese Weise kann man einen schnellen Blick auf die jeweiligen syntagmatischen Verwendungsweisen der gesuchten Phänomene werfen. Wenn die Fundstellen aus Transkripten stammen, die bereits aligniert sind, kann man sich die Fundstelle (in ihrem Kontext) auch unmittelbar von hier aus anhören.

Die entsprechende Recherche für das Suchwort "halt" liefert 9385 Fundstellen aus den Transkripten aller Korpora (Stand: 07/2005). Um mit solch großen Belegzahlen besser umgehen zu können, empfiehlt sich die Aufteilung der Suche nach einzelnen Korpora, die in der Suchmaske entsprechend voreingestellt werden kann. In diesem Fall ist die Kombination der Partikel besonders interessant, also die Frage, ob und in welchen Sprachlandschaften Sätze wie *Das ist halt eben so* bzw. *Das ist eben halt so* vorkommen. Die entsprechende Suchanfrage in der DGD hat eine etwas komplexere Form. Für diese sog. "Abstandssuche", also die Suche nach Wörtern, die im Text in einem bestimmten maximalen Abstand auftreten sollen, wird die Syntax des Rechercheinstruments COSMAS II von der

¹² Siehe auch die Demo-Suchanfrage 2 am Ende der Suchmaske.

DGD-Internetschnittstelle unverändert durchgereicht (in der Suchmaske den Link "Hinweise zur Syntax von Volltextrecherchen" klicken). Nach "halt eben" wird mit der folgenden Vorgabe gesucht: "COSMAS:PROX('halt','+1w', 'eben')", wobei der Abstand zwischen *halt* und *eben* mit "1w", also unmittelbar aufeinander folgend, vorgegeben wird. Diese Suche führt zu 55 Belegen aus unterschiedlichen Sprachlandschaften. Für die entsprechende Suche nach "eben halt" mit dem Suchbefehl "COSMAS:PROX('eben','+1w','halt')", ergeben sich 11 Belege (siehe Abb. 2). Die detaillierte Auswertung der Belege können wir an dieser Stelle nicht liefern. Es sei aber zumindest darauf aufmerksam gemacht, dass die DGD für solche Untersuchungen eine bisher in anderen Forschungszusammenhängen qualitativ und quantitativ nicht annähernd erreichte Fülle von Untersuchungsmöglichkeiten und von Material zur Verfügung stellt.¹³

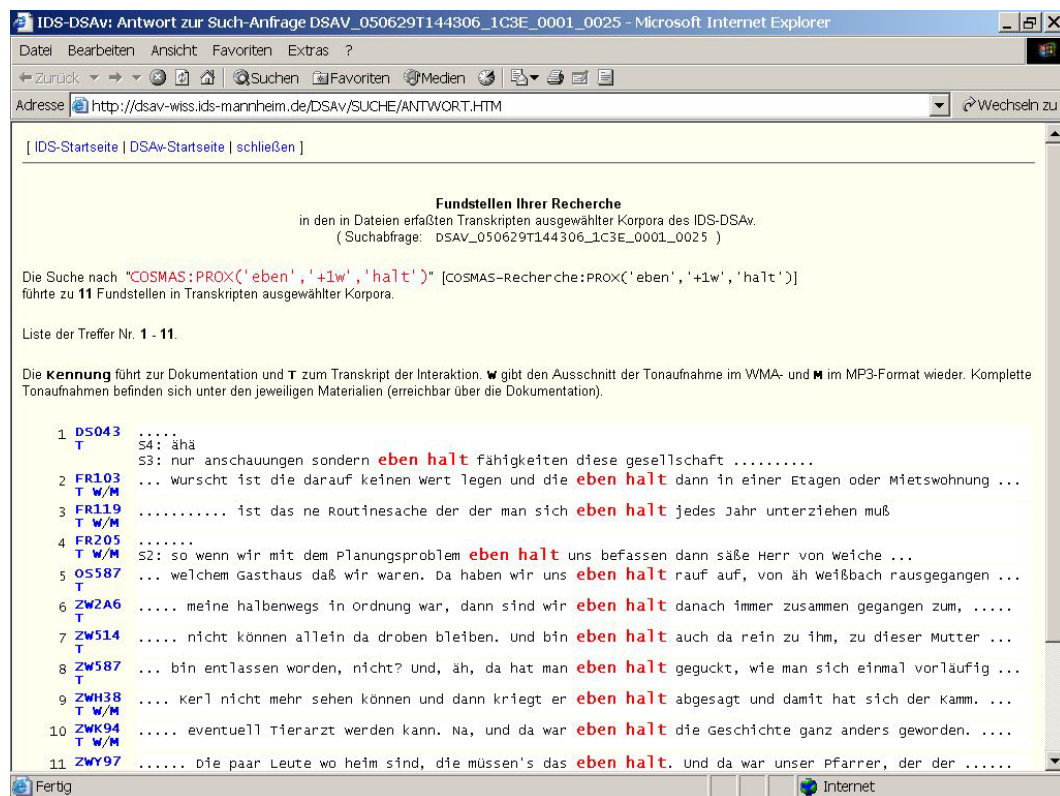


Abb. 2: Liste der Fundstellen zu *eben halt*

¹³ Man vergleiche zum Gebrauch der Partikeln *eben* und *halt* Dittmar (2000). Norbert Dittmar hat aus seinem "Wende"-Korpus die *eben*- und *halt*-Belege für 56 Sprecher(innen) herausgezogen. Die Gesamtzahlen der Belege sind hier 1026 Vorkommen für *eben* gegenüber 365 für *halt*. (Die Tonaufnahmen dieses Korpus wurden im IDS digitalisiert, stehen der DGD aber noch nicht für die Integration zur Verfügung.)

3. Weiterentwicklung und Perspektiven

Da die sehr aufwändige Aufbereitung und Integration der verschiedenen Materialien aus dem DSAv und aus IDS-Forschungsprojekten (angesichts der nicht gerade üppigen Personalausstattung) noch immer im Vordergrund steht, kann die Weiterentwicklung der DGD gegenwärtig nur in kleinen Schritten erfolgen.

Erwähnenswert ist in diesem Zusammenhang die Kooperation von DSAv und DSA, also dem Deutschen Spracharchiv und dem Forschungsinstitut Deutscher Sprachatlas in Marburg. Hintergrund der Zusammenarbeit ist der Gedanke, die weit entwickelte kartografische Kompetenz in Marburg mit der weit entwickelten Tontechnik in Mannheim zu kombinieren. Ein erstes Ergebnis der Kooperation ist die Erweiterung der DGD um eine kartografische Dimension. Dafür wurden die Planquadratangaben, die für viele Orte der DSAv-Aufnahmen dokumentiert wurden (also für Geburts-, Wohn-, Aufnahme-, Schul-, Arbeitsorte u.ä.), als geografische Kodierung ausgewertet und mit Karten verknüpft, die auf dem Server des DiWA (= Digitaler Wenker-Atlas) bereitgehalten werden. In der Wissenschaftlerversion führt jetzt von jeder Planquadratangabe (insgesamt ca. 58 000!) ein Link zu einem Auswahlmenü von vier Karten mit:

- einer Satellitenkarte, die der Identifizierung der geografischen Lage des jeweiligen Ortes dient (zurzeit nur für die Bundesrepublik Deutschland)
- der Karte der Mundartordnungsgruppen des Deutschen Spracharchivs
- der Grundkarte des Deutschen Sprachatlases
- der Karte des Ortsnamensregisters des Deutschen Wortatlases

Je nach Bedürfnis des Nutzers können die Karten der reinen geografischen Identifizierung der Orte oder weiterführenden Informationen in den Karten des DiWA dienen.

Die künftige Weiterentwicklung schließt auch die Erweiterung der Bestände des DSAv ein. Wegen der Dringlichkeit der in den letzten Jahren erfolgten physikalischen Erhaltung der Bestände und ihrer Erschließung durch die DGD sind größere Neuaufnahmeaktionen zur Aktualisierung des Archivs, z.B. zu den gesprochenen Varietäten des Deutschen zwischen Standardsprache und Dialekten oder zu den neuen Regionalsprachen bislang zurückgestellt worden.

In Zukunft wird eine Erweiterung durch die Übernahme externer Korpora verstärkt möglich und nötig sein. Solche gezielten Bestandserweiterungen waren schon immer auch in den Plänen des DSAv verankert. So heißt es schon 1970 in einem "Memorandum zur Weiterführung des Deutschen Spracharchivs" (Ungeheuer/Kohler 1970:4):

Die Konzeption des 'Deutschen Spracharchivs' war von vornherein so angelegt, dass nicht eine einmalige Dialekterhebung in einem begrenzten Zeitraum von Jahren stattfinden würde [...], sondern daß das Projekt ein ständig lebendiges, zeitlich unbegrenztes Forschungsinstrument zur Erfassung phonetischer, syntaktischer, lexikalischer, soziologischer und psychologischer Aspekte des Deutschen in seiner Permanenz und in seinem Wandel über die Jahre hinweg sein sollte.

Geht man diesen Weg konsequent weiter, lassen sich angesichts der erheblich verbesserten technischen Möglichkeiten vorhandene externe Bestände modernisieren, ins DSAv integrieren und auf diese Weise durch ein gut ausgestattetes

zentrales Archiv eine breite und schnell zugängliche Materialbasis für die empirische Sprachforschung etablieren. In diese Richtung zielt auch die oben zitierte Anregung des "Senatsausschuss Evaluierung der Leibniz Gesellschaft". Für die Herstellung "eines ausgewogenen nationalen Referenzkorpus" ist ein zeitgemäßes technologisches Instrumentarium eine unabdingbare Voraussetzung. Die DGD liefert diese notwendige Basis für den Bereich der gesprochenen Sprache weitgehend schon in ihrem heutigen Entwicklungsstand und durch ihre modulare, zukunftsfähige Struktur insbesondere auch im Hinblick auf notwendige, gezielte Erweiterungen.

Literatur

- Dittmar, Norbert (2000): Sozialer Umbruch und Sprachwandel am Beispiel der Modalpartikeln *halt* und *eben* in der Berliner Kommunikationsgemeinschaft nach der 'Wende'. In: Auer, Peter / Hausendorf, Heiko (Hrsg.): Kommunikation in gesellschaftlichen Umbruchsituationen. Mikroanalytische Aspekte des sprachlichen und gesellschaftlichen Wandels in den Neuen Bundesländern. Tübingen: Niemeyer, 199-234.
- Glas, Reinhold / Ehlich, Konrad (2000): Deutsche Transkripte: Ein Repertorium. Arbeiten zur Mehrsprachigkeit, Serie A (63). Hamburg.
<http://www.daf.uni-muenchen.de/DTR/DTR.HTM>
- Hajek, Leo (1937): Das Phonogrammarchiv der Akademie der Wissenschaften in Wien. In: Archiv für die gesamte Phonetik. Erste Abteilung: Archiv für Vergleichende Phonetik 1, 222-230.
- Schmidt, Thomas (2005): Datenarchive für die Gesprächsforschung: Perspektiven, Probleme und Lösungsansätze. In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 6, 103-126.
<http://www.gespraechsforschung-ozs.de/heft2005/heft2005.htm>
- Ungeheuer, Gerold/Kohler, Klaus (1970): Memorandum zur Weiterführung des Deutschen Spracharchivs. Bonn. (Ms.)
- Wagener, Peter/Bausch, Karl-Heinz (1997): Tonaufnahmen des gesprochenen Deutsch. Dokumentation der Bestände von sprachwissenschaftlichen Forschungsprojekten und Archiven. Tübingen: Niemeyer.
- Wagener, Peter (2002): Gesprochenes Deutsch online. Zur Modernisierung des Deutschen Spracharchivs. In: Zeitschrift für Dialektologie und Linguistik 69, 314-335.
- Wagener, Peter (2002): Online Access Tools for Spoken German: The Resources of the Deutsches Spracharchiv in a Database.
http://www.mpi.nl/world/lrec/papers/lrec-pap-17-LREC_Paper.pdf

Prof. Reinhard Fiehler
Dr. Peter Wagener

Institut für Deutsche Sprache
R 5, 6-13
68161 Mannheim

fiehler@ids-mannheim.de
wagener@ids-mannheim.de

Veröffentlicht am 2.11.2005

© Copyright by GESPRÄCHSFORSCHUNG. Alle Rechte vorbehalten.