Erschienen in: Baldauf, Christa J. (Hrsg.): 2. Symposion Autorenerkennung des Bundeskriminalamts vom 03. bis 05. April 2000. Tagungsband. - Wiesbaden: BKA, 2000. S. 1-18.

Instrumente für die Arbeit mit Korpora gesprochener Sprache: Text-Ton-Alignment und COSMAS II

Reinhard Fiehler / Wilfried Schütte, Mannheim

0 Einleitung

Straftaten werden nicht nur durch praktische Handlungen begangen (wie z.B. beim Diebstahl oder der Körperverletzung), sondern zu einem wachsenden Anteil bestehen sie in kommunikativen Aktivitäten, werden sie kommunikativ ausgeführt – und dies sowohl in schriftlicher wie auch in mündlicher Form. Betrachten wir – worauf wir uns in unseren Ausführungen beschränken möchten – die mündliche Form, so kann das gesprochene Wort selbst die Straftat darstellen (wie z.B. bei Beleidigungen), oder es kann dazu benutzt werden, um Straftaten vorzubereiten und auszuführen (wie z.B. bei Gesprächen zur Planung von Straftaten, bei Fällen von Nötigung oder bei erpresserischen Anrufen). Es kann aber auch eingesetzt werden, um Straftaten zu verhindern, nicht eskalieren zu lassen oder zu beenden (wie z.B. bei Verhandlungen mit Straftätern im Vollzug der Tat).
Als Folge dieser Bedeutsamkeit des gesprochenen Worts im Kontext von Straftaten werden in der kriminalistischen Arbeit immer mehr Gespräche mitgeschnitten, bearbeitet und ausgewertet. Dies reicht von Lauschangriffen auf verdächtige Personen über die Aufzeichnung krimineller Anrufe (Erpressungen, Drohanrufe, sexuelle Belästigungen, Falschmeldungen bei Notdiensten etc.) bis hin zu Mitschnitten der Verhandlungsführung mit Geiselnehmern, Suizidkandidaten etc. Die Auswertung dieser Gesprächsaufzeichnungen kann dabei auf ganz unterschiedliche Ziele gerichtet sein: Die Vereitelung von Straftaten, die Identifizierung von Straftätern (Stimmerkennung), die Optimierung von Verhandlungsstrategien[1] etc.
Wir gehen also davon aus, dass die Dokumentation und Analyse straftatbezogener Sprachaufzeichnungen[2] zumindest ebenso wichtig ist wie die entsprechende Auswertung schriftlicher Texte. Während aber die Dokumentation und Analyse schriftlicher Texte – sowohl in der Linguistik wie auch in der Kriminologie (als Folge der gesellschaftlichen Dominanz der geschriebenen Sprache) – einen hohen Stand erreicht haben, besteht im Bereich der gesprochenen Sprache hier ein deutlicher Nachholbedarf. Dies liegt u.a. auch an den besonderen Schwierigkeiten des Umgangs mit gesprochener Sprache: Die Dokumentation und Auswertung von Sprachdaten bringt – wie jeder weiß, der damit beschäftigt ist – erheblich größere Probleme mit sich, als es bei schriftlichen Texten der Fall ist. Während schriftliche Texte unmittelbar 'vor

[1] Vgl. Brünner/Fiehler/Wiegers 1996.
[2] Wenn uns diese Anmerkung gestattet ist: Aus linguistischer Perspektive halten wir es für wünschenswert, im Kontext einer Straftat (z.B. einer Erpressung) die Dokumentation von kommunikativen Ereignissen nicht auf unmittelbare Täteräußerungen zu beschränken, sondern im Sinne einer Ethnographie der Ermittlung das kommunikative Geschehen sehr viel umfangreicher zu dokumentieren (Aussagen von anderen Beteiligten, Zeugen etc., Besprechungen im Rahmen der Ermittlungen usw.). U.E. ist dies nicht nur gesprächsanalytisch interessant, sondern kann auch für die Ermittlungen selbst von Bedeutung sein (wer hat was wann genau gesagt). Die übliche Vorgehensweise scheint uns zu 'täterzentriert'.

Augen stehen' und eine Orientierung in ihnen relativ einfach ist, müssen Gesprächsaufzeichnungen zeitaufwendig abgespielt werden, um sie sich zu vergegenwärtigen und um sich in ihnen zu orientieren. Besondere Probleme bereiten beim Umgang mit Sprachdaten z.B.

- die Dokumentation und Archivierung der Aufnahmen,
- die langfristige Aufbewahrung und der Erhalt von Aufnahmen,
- das Auffinden bestimmter Stellen in Aufnahmen,
- das Ermitteln von Vergleichsstellen (innerhalb einer oder verschiedener Aufnahmen) und die Durchführung entsprechender Vergleiche.

Vor den hier beschriebenen Problemen steht aber nicht nur die Kriminalistik, sondern auch die linguistische Gesprächsforschung: Sie sind sozusagen ihr 'täglich Brot'. Die Gesprächsforschung hat sich in den letzten 25 Jahren als eigenständige Teildisziplin innerhalb der Sprachwissenschaft und Soziologie etabliert. Ihr Ziel ist die wissenschaftliche Erforschung der Organisationsprinzipien von mündlicher Kommunikation und der Regularitäten des kommunikativen Handelns in Gesprächen. Die *Aufzeichnung* authentischer Gespräche, ihre detaillierte *Verschriftlichung* (Transkription) und die *Analyse* dieser Transkripte und Aufnahmen unter aus dem Material entwickelten Fragestellungen bilden den Kern der gesprächsanalytischen Arbeitsweise (für Näheres vgl. Becker-Mrotzek/Meier 1999; Deppermann 1999).
Zur Erleichterung ihrer Arbeit hat die Gesprächsforschung eine Reihe von Hilfsmitteln entwickelt. An erster Stelle zu nennen ist die schon erwähnte Transkription von Aufnahmen, bei der das gesprochene Wort mit seinen prosodischen Eigenschaften in einer relativ feinen Weise in einen schriftlichen Text umgesetzt wird (s.u. Abschnitt 2). Die Analyse erfolgt dann auf der Grundlage dieser Transkripte und der dazugehörenden Aufnahmen.
Wenn nicht nur einzelne Gesprächsaufnahmen zu Transkripten verarbeitet worden sind, sondern viele derartige Aufnahmen nach Varianz- oder Ähnlichkeitsgesichtspunkten zusammengestellt werden, sprechen wir von Korpora. Gesprächskorpora können u.a. nach systematischen Gesichtspunkten entstehen, etwa um Gespräche eines bestimmten Interaktionstyps zu dokumentieren (z.B. „Beratungen" oder „Erpresseranrufe") oder um Tendenzen der Sprachentwicklung aufzuzeigen, indem Gespräche nach ihrem Aufzeichnungsdatum chronologisch zusammengestellt werden. Solche Korpora können zu einem Archiv zusammengefügt werden.[3]
Mit der Transkription und der Archivierung von Transkripten in Korpora entstehen weitere Anforderungen, nämlich:

- zu einem Transkriptstück den entsprechenden Ausschnitt in der Aufnahme zu finden und anhören zu können,
- alle Vorkommen eines Phänomens in den Transkripten aufzufinden (und parallel dazu die entsprechenden Tonausschnitte hören zu können),
- nicht nur in einem Transkript, sondern im archivierten Korpusbestand oder in passend zur aktuellen Fragestellung zusammengestellten Teilkorpora recherchieren zu können.

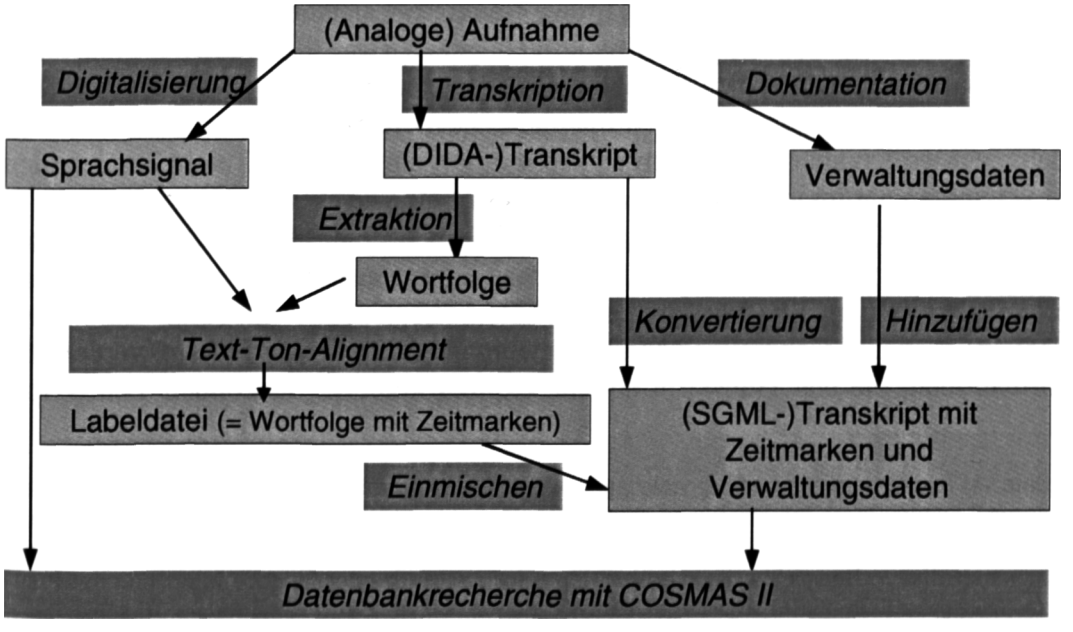
---

[3] So verwaltet das Deutsche Spracharchiv (DSAv) am Institut für Deutsche Sprache die 33 gesprochensprachlichen Korpora des Instituts: vgl. http://www.ids-mannheim.de/dsav/

Zur Bearbeitung dieser drei Anforderungen sind in den letzten Jahren in der Abteilung 'Pragmatik' des Instituts für Deutsche Sprache (IDS) in Mannheim, in der vor allen gesprächsanalytische und soziolinguistische Untersuchungen durchgeführt werden, Arbeitsinstrumente entwickelt worden, die im Folgenden vorgestellt werden sollen. Das automatische Text-Ton-Alignment und das Recherchesystem COSMAS II erleichtern die Arbeit mit einzelnen Aufnahmen wie auch mit ganzen Korpora gesprochener Sprache erheblich. Ermöglicht wurden diese Entwicklungen durch den Übergang vom analogen zum digitalen Tonsignal und durch die Übertragung der Korpusarbeit ins elektronische Medium.
Sofern also die Kriminalistik und die Gesprächsforschung – was den Umgang mit dem gesprochenen Wort betrifft – strukturell vor den gleichen Aufgaben und Problemen stehen, haben wir die Hoffnung, dass die genannten Arbeitsinstrumente auch im kriminalistischen Bereich einen sinnvollen Einsatz finden und die Arbeit erleichtern können.
Bevor wir in Abschnitt 3 das Text-Ton-Alignment und in Abschnitt 4 COSMAS II vorstellen, soll zum besseren Verständnis, wie diese Instrumente eingebunden sind, in Abschnitt 1 ein Überblick über die Abläufe bei der Bearbeitung von Tonaufzeichnungen gegeben und in Abschnitt 2 das Transkriptionsverfahren beschrieben werden.

1 Arbeit mit Gesprächsaufnahmen und -korpora

Die folgende Grafik stellt in Grundzügen den Ablauf dar, wie am IDS Gesprächsaufnahmen verarbeitet werden, so dass sie schließlich über eine Datenbank recherchierbar werden. Im Prinzip sind diese Arbeitsschritte aber für jedes Gesprächskorpus notwendig, gleichgültig in welcher Zusammensetzung und in welcher Notationsform, wenn man ein Text-Ton-Alignment durchführen und die Recherche in einer Datenbank ermöglichen will.



Endziel der Bearbeitung von Gesprächsaufnahmen ist die **Recherche in der Gesprächsdatenbank** mit COSMAS II. Dabei sollen Treffer sowohl im Transkript angezeigt werden als auch im Ton abgespielt werden können. Um dieses Ziel zu erreichen, sind folgende Arbeitsschritte (= blaue Markierungen) notwendig, die Zwischenprodukte (= gelbe Markierungen) ergeben:

- **„(Analoge) Aufnahme":** Die Grundlage bilden Aufnahmen von natürlichen Gesprächen, die – zumindest traditionell – auf Audio-, gelegentlich auch auf Videocassetten aufgenommen wurden. Natürlich ist die technische Qualität dieser Aufnahmen eingeschränkt und nicht mit Studioqualität zu vergleichen: Die älteren Aufnahmen sind zumeist nur in mono (was beim Transkribieren die Ortung und damit die Zuordnung von Redebeiträgen zu Sprechern erschwert), wurden mit möglichst kleinen Aufnahmegeräten gemacht (im Bestreben, die Gesprächssituation möglichst wenig zu stören) und weisen Hintergrundgeräusche auf.
- „**Transkription**": Gespräche bzw. ihre Tonaufzeichnung sind für die empirische Gesprächsforschung die primären Daten; Untersuchungen dieser Gespräche ließen sich – wiederum traditionell gesehen – aber nur durchführen und intersubjektiv nachprüfbar machen, indem die Gespräche nach einem standardisierten Transkriptionssystem verschriftet worden waren. Am IDS wurden dazu ein Editor namens „DIDA" (= DIskurs-DAtenverarbeitung), mit dem man Gespräche in Partiturschreibweise aufzeichnen kann, und ein zugehöriges Notationssystem entwickelt, dessen Merkmale im folgenden Abschnitt beschrieben werden.
- „**Dokumentation**": Zu den Gesprächen, den Aufnahmen und den Transkripten wurden Verwaltungsdaten erhoben, die z.B. auf Transkript-Deckblättern festgehalten werden: Aufnahmedaten, Datenträger, Sprecher (und zugehörige Siglen in Transkripten), Besonderheiten, Kommunikationsverlauf.
- Für eine computergestützte Verarbeitung der Gesprächsaufnahmen muss ein **(digitales) Sprachsignal** vorliegen; die alten analogen Feldaufnahmen müssen also zunächst digitalisiert werden. Dabei greifen wir aus archivarischen Gründen zu relativ großzügigen Parametern (WAV-Dateien mit 48 kHz Abtastrate in 16bit-Quantisierung, je nach analoger Vorlage mono oder stereo; die nach Nyquist dadurch erfassbaren Sprachsignale bis 24 kHz gehen zwar über die Obergrenze des durchschnittlichen menschlichen Gehörs von je nach Alter 15-20 kHz und auch über das, was mit älteren analogen Aufnahmegeräten aufgezeichnet werden konnte, hinaus, wir wollen aber auch für prosodische Analysen das Signal nicht beschneiden und verzichten darum auch auf Kompressionsformate wie etwa MP3).
- Die Datenbank COSMAS II greift auf dieses Sprachsignal und auf die Transkripte zurück, allerdings müssen die Transkripte dazu in ein besonderes Format konvertiert werden, nämlich in **TEI-konformes SGML** (TEI = „Text Encoding Initiative"; SGML = „Standard Generalized Markup Language"). In diesem Format werden alle Informationen, die das Transkript enthält, explizit kodiert.[4] So können Transkripte

[4] Transkripte auf Papier informieren über viele signifikante Vorgänge mit ikonischen Zeichen oder mittels des Transkript-Layouts; diese Informationen sind nur im Zusammenhang mit einer Legende verständlich und trotz nachhaltig verfolgter Initiativen (vgl. Selting u.a. 1998) in der gesprächsanalytischen Forschung noch nicht standardisiert. Beispielsweise wird in DIDA eine steigende oder fallende Intonation mit Pfeil-Zeichen (↑ bzw. ↓) notiert; Simultanpassagen (Überlappungen zwischen Redebeiträgen, Sprecher reden gleichzeitig) werden analog zu musikalischen Partituren durch übereinander stehende Äußerungsteile wiedergegeben. In unseren SGML-Transkripten werden die steigende Intonation mit dem Tag „<shift feature=intonation here=steigend>„ die fallende Intonation mit dem Tag „<shift feature=intonation

aus einem beliebigen proprietären Format über Betriebssystem- bzw. Plattformgrenzen und einzelne Anwendungen, insbesondere Transkript-Editoren, hinaus in ein weit verbreitetes und auch zukunftssicheres Austauschformat gebracht werden. Das SGML-Transkript soll auch zu jedem Wort die Zeitmarken und zum Zweck einer Korpusauswahl zum gesamten Transkript die Verwaltungsdaten enthalten.

- Um die Zeitmarken zu gewinnen, muss nun ein **Text-Ton-Alignment** durchgeführt werden; Input für dieses Verfahren sind eine aus dem komplexen Transkript extrahierte einfache Wortfolge und das Sprachsignal (dafür ist aus Speicherplatz- und Performance-Gründen eine auf eine Abtastrate von 16 kHz reduzierte „Arbeitskopie" des Sprachsignals sinnvoll). Das Text-Ton-Alignment ergibt eine sogenannte „Labeldatei": eine um Zeitmarken angereicherte Wortfolge.
- Die Konvertierung besorgt ein im IDS entwickeltes Programm, das in einem Arbeitsschritt drei Aufgaben erledigt:
  (a) Das DIDA-Transkript wird aus einem proprietären Mailbox-Format, das zur Kommunikation zwischen DIDA-Server und -Client dient, in TEI-konformes SGML konvertiert;
  (b) die aus dem Alignment gewonnenen Zeitmarken werden den Wörtern des Transkripts zugeordnet, sozusagen in die SGML-Datei „eingemischt";
  (c) die Verwaltungsdaten werden hinzugefügt.

Die Gesprächskorpora am IDS umfassen u.a. Gespräche aus mehreren Kommunikationsdomänen , teils aus institutionellen Zusammenhängen, nämlich

- Beratungsgespräche unterschiedlicher Art;
- Schlichtungsgespräche (z.B. aus der außergerichtlichen Schlichtung von Nachbarschaftsstreitigkeiten vor einer Vergleichsbehörde oder aus Verfahren, mit denen Verbraucherreklamationen vor Schlichtungsstellen von Handwerkskammern behandelt werden);
- Aufnahmen aus dem Projekt „Stadtsprache Mannheim", in dem Formen sowie kommunikative und soziale Funktionen des „Monnemerischen" untersucht wurden;
- Gespräche im Fernsehen (also Talkshows und Diskussionen);
- Ethnografische Interviews mit Fernsehredakteuren und -moderatoren;
- Interviews mit Beamten der europäischen Institutionen in Brüssel, den sog. „Eurokraten".

Diese natürlich noch nicht für alle Kommunikationsbereiche des gesprochenen Deutsch repräsentativen Korpora werden fortlaufend erweitert. Eine vollständige Auflistung des gegenwärtigen (z.T. auch für Servicezwecke zur Verfügung stehenden) Bestands findet man auf dem WWW-Server des IDS.[5]

---

here=fallend>„ und Anfang und Ende von Simultanpassagen in allen beteiligten Äußerungen mit Tags wie „<anchor n=2 id=simCC2B type=sim function=beg>„ bzw. „<anchor n=2 id=simCC2E type=sim function=end>„ kodiert.

[5] Vgl. http://www.ids-mannheim.de/dsav/korpora/korpusliste.html

2 Transkription

An einem Ausschnitt aus einem Schlichtungsgespräch möchten wir illustrieren, wie am IDS transkribiert wird. Die Verschriftlichung der Gespräche in DIDA basiert auf folgenden Prinzipien:

- **Literarische Umschrift**: Sie stützt sich grundsätzlich auf das orthografische System und ergänzt dieses zur Präzisierung der Lautwiedergabe um eine Reihe von Sonderzeichen. Dadurch soll auch die im Einzelfall besondere Artikulation, z.B. dialektaler oder umgangssprachlicher Art, der Sprecher im Transkript wiedergegeben werden. Die Umschrift orientiert sich grundsätzlich an den Regeln der Standard-Orthographie, verzichtet aber auf die Großschreibung, die normale Interpunktion sowie die Trennung am Zeilenende.[6]
- DIDA-Transkripte werden grundsätzlich in **Partiturschreibweise** angefertigt, d.h. für jeden am Diskurs beteiligten Sprecher existiert eine eigene Zeile, auf der seine Äußerungen verschriftlicht werden. Die Reihenfolge der Sprecher innerhalb des Partiturblocks bleibt über das gesamte Transkript hinweg konstant. In den Transkripteditor von DIDA wird fortlaufend auf einer endlosen Partiturzeile eingegeben, metaphorisch gesprochen: wie auf einer „Papierrolle".
- Die **prosodische Notation** umfasst Grenzintonationsmuster an Stellen möglicher Redeübergabe (also möglichen Sprecherwechsels), Pausen, Wechsel in der Sprechweise (Lautstärke und Sprechgeschwindigkeit).
- **Interaktivität:** Weil wir die Äußerungen sprecherbezogen als „Turns" im Partiturformat, also im Sprecherzeilenblock, notieren, können Simultanpassagen (Überlappungen, Einwürfe oder Rezeptionssignale) berücksichtigt werden.
- **Kommentare:** Zu jedem Sprecher sowie zu Segmenten des gesamten Partiturblocks sind Kommentarzeilen möglich. Darin finden sich analytische Beschreibungen mit lokaler Referenz zur Sprechweise oder zu interaktionsrelevanten nichtsprachlichen Aktivitäten einzelner Sprecher oder global zum Diskurs.
- Personenbezogene Daten werden **anonymisiert bzw. maskiert** (das gilt für Personen-, Ortsnamen, Aktenzeichen u.ä.).

Die nächste Grafik zeigt ein Bildschirmfoto vom DIDA-Transkripteditor, nämlich ein Beispiel aus einem Schlichtungsgespräch mit der Verwaltungsnummer 3001.03,altesau.

[6] Für die Datenbank-Recherche ergibt sich hier ein Dilemma: Je genauer die Artikulation im Einzelfall als Abweichung von der standardsprachlichen Orthografie notiert wird, desto brauchbarer ist das Transkript, um Muster für umgangssprachliche Formulierungen zu belegen, desto zufälliger wird aber auch das Ergebnis der Datenbankrecherche. Anfragen müssten alle denkbaren Verschriftungsformen angeben, die man aber vor der Recherche nicht kennt. Der Ausweg ist eine Lemmatisierung, also eine regelbasierte automatische Zuordnung aller Flexionsformen, umgangssprachlichen und dialektalen Varianten zu einer Grundform, einem „Lemma". Diese Lemmatisierung ist für die Zukunft geplant.

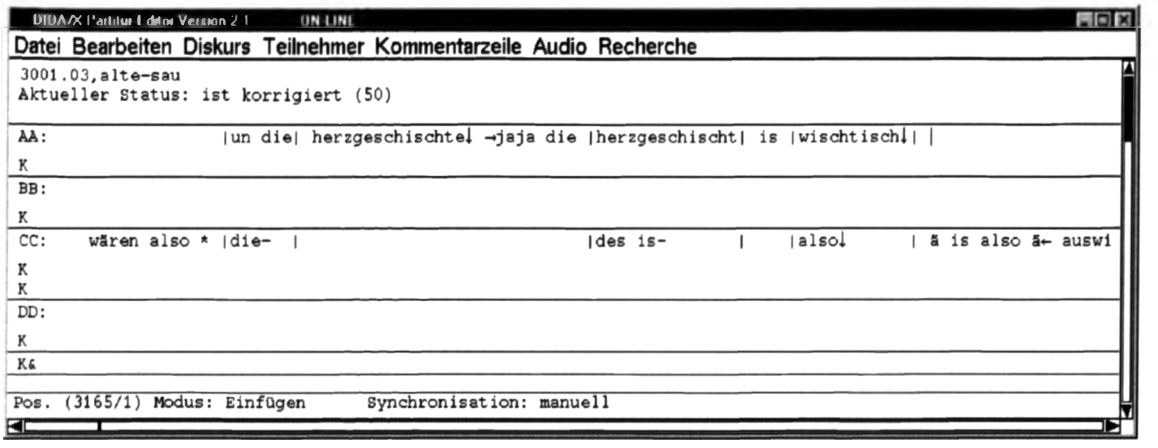


Abb. Bildschirmfoto des DIDA-Transkripteditors mit einem Ausschnitt aus einem Schlichtungsgespräch

Erkennbar sind hier die Sprecherzuordnung (AA und BB sind die Kontrahentinnen, wobei BB in diesem Ausschnitt nicht redet, CC ist der Schlichter), drei Simultanpassagen (AA und CC reden gleichzeitig; Anfang und Ende dieser Passage sind in beiden Sprecherzeilen durch | markiert), eine rudimentäre prosodische Notation (mit Grenzintonationsmustern, markiert durch ↑ und ↓; Veränderung der Sprechgeschwindigkeit, markiert durch →; Pausen, markiert durch *), „literarische Umschrift" (standardisierte Dialektnotation, z.B. bei „wischtisch" und „des is"). Nähere Informationen zu diesem Editor findet man im IDS-WWW-Angebot.[7]
Transkripte können aus dem Editor heraus in ein Textformat exportiert werden, damit sie als Belege für analytische Texte zur Verfügung stehen und in diese eingefügt werden können. Dabei bleibt die Darstellung im Partiturformat erhalten. Sprecherzeilen werden durchnummeriert, Anfang und Ende von Simultanpassagen wiederum durch senkrechte Striche (|) gekennzeichnet:

```
[...]
66   CC:          mhm ja donn les isch=s ihne noch vor damit

67   CC: sie=s au"ch äh gehört haben du drecksau du wildsau-

68   CC: * geh in=s aldersheim- * wenn keine kinder leiden

69   CC: kannst du alder schrubber geh runder oder isch zieh

70   CC: disch an den haaren herbei- * äh du dreckeber mit

71   CC: deinem bappalten↓ du gehörst vergast↓ ** des wären

72   AA:         |un die| herzgeschischte↓ →jaja die
73   CC: also * |die-   |

74   AA: |herzgeschischt| is |wischtisch↓|
75   CC: |des is-       |    |also↓      | ä is also ä←

76   AA:                 |ja is für misch seh"r wischtisch↓|
```

[7] Der DIDA-Transkripteditor ist Teil eines Systems „DIDA-Diskursdatenbank"; vgl. http://www.ids-mannheim.de/prag/dida/.

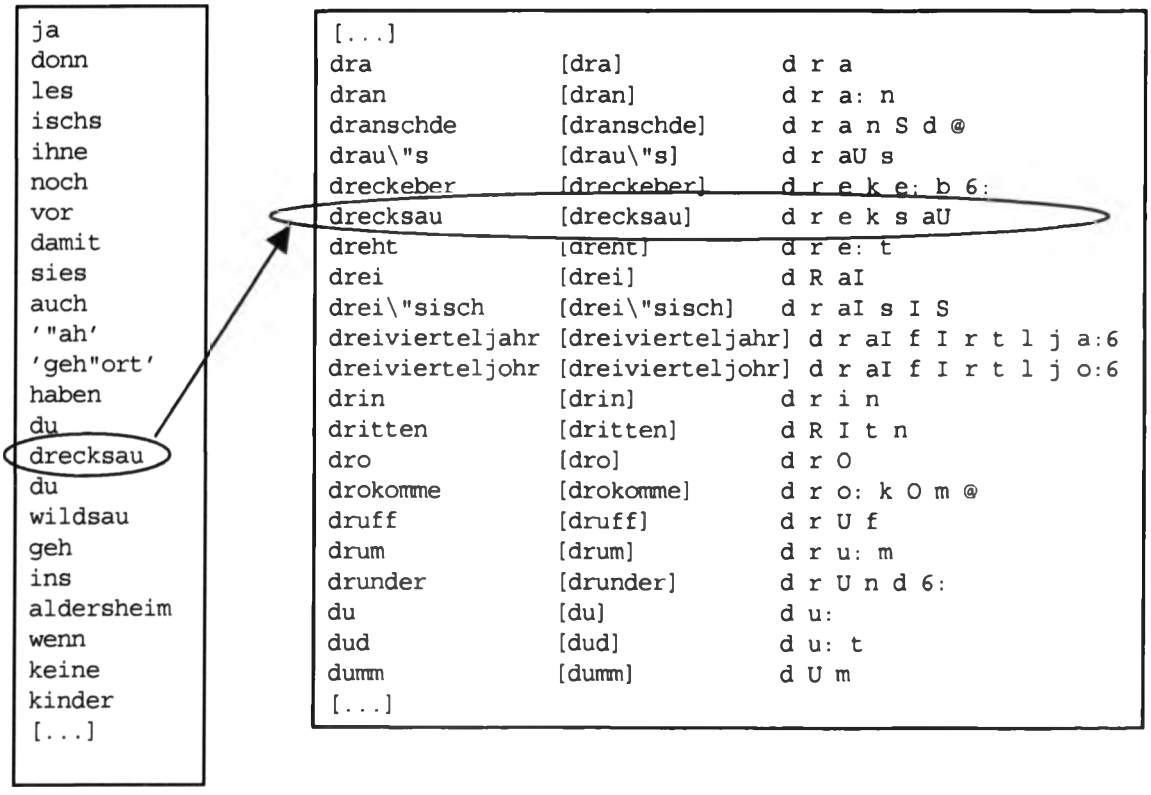
```
77   CC: auswirkung der |sache net is also die frau die    |

78   CC: frau beck gibt also hier an sie hätte also zwei

9   CC: näschte lang furschbare herzkrämpfe gehabt un sie
79 [...]
```

4 Text-Ton-Alignment

Beim Text-Ton-Alignment werden das Transkript und das Sprachsignal durch eine Zuordnung von Zeitmarken zu jedem Wort des Transkripts synchronisiert. Hinsichtlich des Transkripts sind dabei zwei Vorbereitungsschritte notwendig: (a) die Extraktion einer Wortfolge, bei der alle Zusatzinformationen (z.B. Sprecherzuordnung, Prosodie, Pausen) aus dem Transkript zunächst ausgeblendet werden, und (b) eine regelbasierte automatische Phonetisierung, bei der die orthografische Notation in eine phonetische SAMPA-Kodierung überführt wird (SAMPA ist eine IPA-ähnliche Notation, die mit den Zeichen des ASCII-Codes dargestellt werden kann) und die Einträge alphanumerisch sortiert werden. Als Beispiel hier das Schlüsselwort aus dem Schlichtungsgespräch, das eben ausschnittweise im exportierten Transkript gezeigt wurde, nämlich die Invektive „drecksau“:



An dieser Stelle sei darauf verzichtet, die technischen Einzelheiten des automatischen Alignments darzustellen. Das Verfahren basiert auf Hidden-Markov-Modellen und ist zusammengestellt aus den Bausteinen, die der HTK-Toolkit der englisch-

amerikanischen Fa. Entropic bietet. Die Aufgabenstellung beim Alignment ist erheblich einfacher als bei der automatischen Spracherkennung. Beim Alignment werden dem Signal nicht unbekannte Wörter zugeordnet, sondern es wird lediglich für ein bekanntes Wort (in einer feststehenden Folge von Wörtern) die beste Entsprechung im Tonsignal gesucht und das Zeitintervall festgehalten.[8] Das Ergebnis des Text-Ton-Alignments ist dann eine sog. „xlabel"-Datei, in der tabellenartig den einzelnen Wörtern der Wortfolge Zeitmarken zugeordnet sind:

```
[...]
214.96 121 ja
215.10 121 donn
215.38 121 les
215.52 121 ischs
215.75 121 ihne
216.12 121 noch
216.19 121 vor
217.57 121 damit
217.66 121 sies
217.88 121 auch
217.98 121 '"ah'
218.20 121 'geh"ort'
218.65 121 haben
218.71 121 du
219.17 121 drecksau
219.24 121 du
219.88 121 wildsau
220.39 121 geh
220.65 121 ins
221.46 121 aldersheim
222.45 121 wenn
222.83 121 keine
```

```
223.13 121 kinder
223.53 121 leiden
224.07 121 kannst
224.22 121 du
224.54 121 alder
225.11 121 schrubber
225.24 121 geh
225.74 121 runder
226.05 121 oder
226.24 121 isch
226.51 121 zieh
226.65 121 disch
226.83 121 an
226.94 121 den
227.37 121 haaren
227.91 121 herbei
229.14 121 '"ah'
229.20 121 du
230.03 121 dreckeber
230.17 121 mit
230.82 121 deinem
[...]
```

Die Ergebnisse des automatischen Alignments lassen sich u.a. visualisieren, anwenden und weiterverarbeiten, indem Labeldateien und Sprachsignale mit einem speziellen Prosodie-Programm, nämlich „Praat" eingelesen werden. „Praat" ist ein Programm zur computergestützten Analyse von Sprachsignalen.[9] „Praat" läuft unter allen relevanten Betriebssystemen bzw. auf allen relevanten Plattformen (d.h. verschiedene Unix-Dialekte, Linux, MacOS, Windows 95/98, NT und 2000).
Die Übernahme von Sound- und Labeldateien nach „Praat" lässt sich schematisch so darstellen:

[8] Das Text-Ton-Alignment lässt sich auch als „simplified speech recognition" beschreiben (vgl. Rapp 1995).
[9] Der Autor ist Paul Boersma, Universität Amsterdam. E-Mail: paul.boersma@hum.uva.nl; http://www.praat.org.

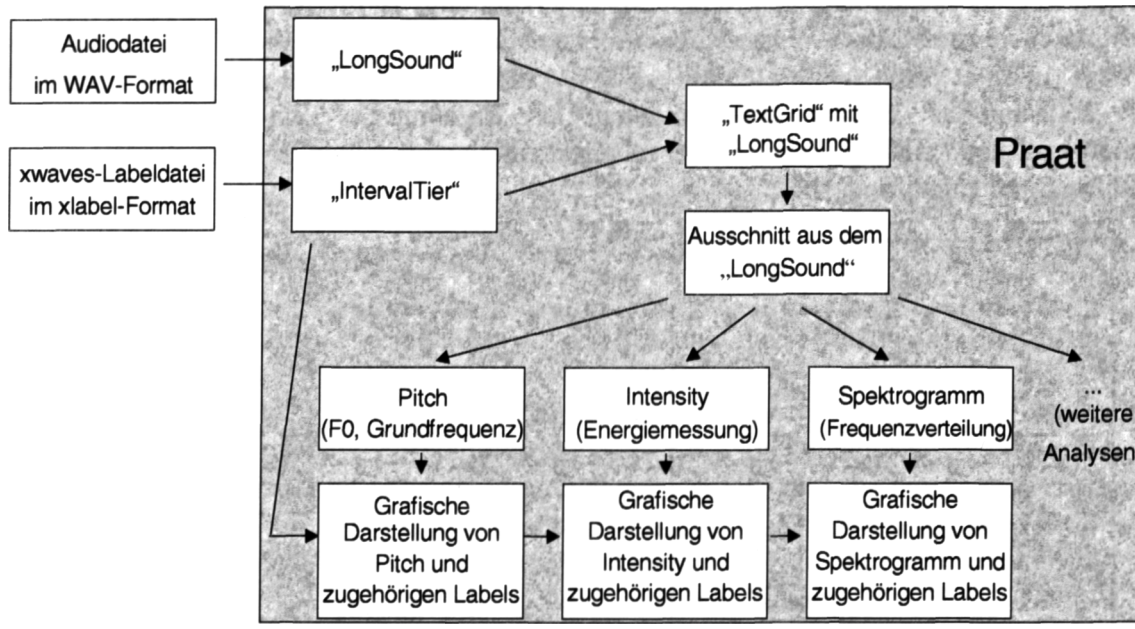


Abb.: Übernahme von Sound- und Labeldateien nach „Praat"

- Ausgangspunkt sind Audiodateien in einem gängigen Format, z.B. WAVE-Dateien, und Labeldateien;
- Sie werden in Praat paarweise eingelesen als sog. „LongSound" und als „IntervalTier", also Annotationsdateien.
- Diese werden in einem sog. „TextGrid-Editor" gemeinsam dargestellt.
- Über eine Suche in der Labeldatei oder durch Abhören der Audiodatei kann man analyserelevante Ausschnitte aus dem Sprachsignal bestimmen.
- Diese lassen sich nun hinsichtlich verschiedener Parameter analysieren. So kann man die Intonationskurve durch Grundfrequenzanalysen (f0-Kurve), die Akzentuierung durch Kombination aus Grundfrequenzkurven, Pausensetzung und Intensität (energy) sowie die Vokalqualität durch Spektrogramme (Verteilung hinsichtlich Zeit, Frequenz und Intensität) bestimmen.
- Alle Analysen lassen sich auch grafisch darstellen; dabei sind die importierten Wort-Label bei der Identifizierung und Zuordnung von Segmenten aus dem bestimmen Sprachsignal, aus der Grundfrequenz-Kurve, der Intensitäts-Kurve und dem Spektrogramm eine große Hilfe und Arbeitserleichterung.

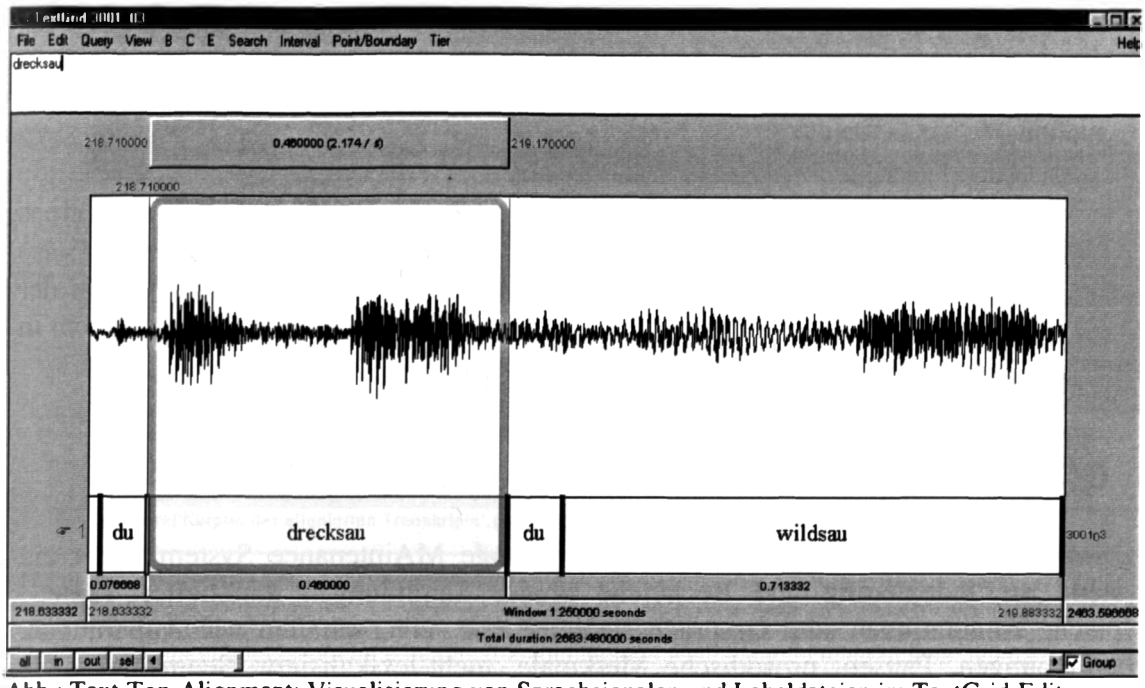


Abb.: Text-Ton-Alignment: Visualisierung von Sprachsignalen und Labeldateien im TextGrid-Editor von „Praat"

Die Ergebnisse des Text-Ton-Alignments lassen sich auch mit einem anderen Programm darstellen, nämlich dem französischen „Transcriber"; hier laufen – fast wie in einem Zeichentrickfilm – der Cursor durch das Sprachsignal und eine Markierung synchron dazu durch die Wortfolge. „Transcriber" eignet sich allerdings nicht zum Segmentieren des Sprachsignals und zur weiteren prosodischen Analyse.[10]
Programme wie „Praat" bieten reichhaltige Möglichkeiten für eine textbasierte Recherche, haben dabei aber auch ihre Grenzen. Die Möglichkeiten sind (angenommen, die Audiodatei sei korrekt und vollständig aligniert worden):

- ein schnelles Aufsuchen von Stellen im Sprachsignal, denen ein bestimmtes Wortlabel zugeordnet wurde („nicht-sequenzieller Zugriff");
- diese Ausschnitte aus dem Sprachsignal können für die Analyse (z.B. Prosodie, Spektrogramm) weiterverwendet werden;
- ebenso können die aus dem Alignment übernommenen Wortlabels für die Analyse und grafische Präsentation weiterverwendet werden.

Aber es gibt auch Grenzen, die andere Lösungen erforderlich machen. „Praat" ist keine Datenbank, weswegen es z.B. starke Einschränkungen bei den Recherchemöglichkeiten gibt:

- Es ist nur eine Suche nach gelabelten Wörtern möglich;
- Zusatzinformationen aus dem Transkript (u.a. Sprecherzuordnungen, Kommentare und prosodische Markierungen) werden vernachlässigt;

[10] Vgl. http://www.etca.fr/CTA/gip/Projets/Transcriber/: „Transcriber is a tool for assisting the creation of speech corpora. It allows to manually segment, label and transcribe speech signals, for later use in automatic speech processing. It is more specifically geared towards the transcription of long duration broadcast news recordings, with labeling of speech turns and topic changes. It provides a user-friendly interface which is configurable."

- bei Simultanpassagen ist die aus dem Transkript extrahierte Wortfolge auf den Turn des etablierten Sprechers beschränkt, daher kann man auch nicht ohne Weiteres nach den Äußerungen oder Äußerungssegmenten von hinzukommenden Sprechern suchen;
- eine präzise Übernahme der Label-Schreibweise ist notwendig;
- es gibt keine Recherchesyntax wie in COSMAS II (z.B. unter Benutzung eines Grundformenoperators oder der Definition von Wortabständen);
- eine Suche und der Treffer-Zugriff auf das zugehörige Sprachsignal ist nur in der aktuell geladenen Label-Datei möglich, nicht im gesamten Korpus und auch nicht in beliebig zusammengestellten Teilkorpora.

5 COSMAS II

COSMAS II – ein Akronym aus „COrpus Storage MAintenance System" - ist ein System zur Indexierung und Recherche großer Sammlungen annotierter Texte.[11] Typische Annotationen sind Gesprächselemente und -eigenschaften wie Äußerungen, Überlappungen, Pausen, prosodische Merkmale, nicht-lexikalisierte Segmente sowie (sprecherbezogene oder globale) Kommentare. Bei Recherchen werden diese Besonderheiten verschrifteter gesprochener Sprache berücksichtigt; es gibt zudem einen sprecherbezogenen Wortabstandsoperator.

- Beliebige Teilkorpora können zusammengestellt werden, z.B. nach der Transkriptliste oder nach anderen Verwaltungsdaten;
- die SGML-Transkripte werden auf einem Server indexiert;
- Suchanfragen werden über eine grafische Abfragesprache formuliert; die Anfragekomponente stellt eine Reihe von „primitiven", graphisch verschiebbaren Suchoperatoren zur Verfügung, die gemäß der Syntax der Anfragesprache untereinander oder iterativ mit bestehenden Resultaten kombinierbar sind. Formulierte Suchanfragen können ihrerseits mit einem Namen versehen und zurückgelegt werden, damit man sie später wiederverwenden kann;
- die Anzeige von Recherchetreffern ist im Transkript und im Sprachsignal möglich;
- Sprachsignal-Ausschnitte können an ein Analyseprogramm wie „Praat" übergeben werden.

Die Arbeit mit COSMAS II geht in folgenden Schritten vor sich:

- **Korpusauswahl**: Zunächst wird aus der Suchpalette ein Korpus für die Suche bestimmt, z.B. das Gesamtkorpus der Transkripte (das alle indexierten DIDA-Transkripte umfasst), das Korpus der alignierten Transkripte (das derzeit nur eine Auswahl aus allen Transkripten bietet) oder nach bestimmten Selektionskriterien; so kann man über die Datierung (Aufnahmedatum) zwei Korpora bestimmen, etwa um zu Fragen der Sprachentwicklung (z.B. Neologismen) zu recherchieren. Nach einer Suche im Dokumentenbestand wird das Ergebnis rechts oben in der Ergebnisliste dargestellt.

[11] Vgl. http://www.ids-mannheim.de/zdv/cosmas2/.

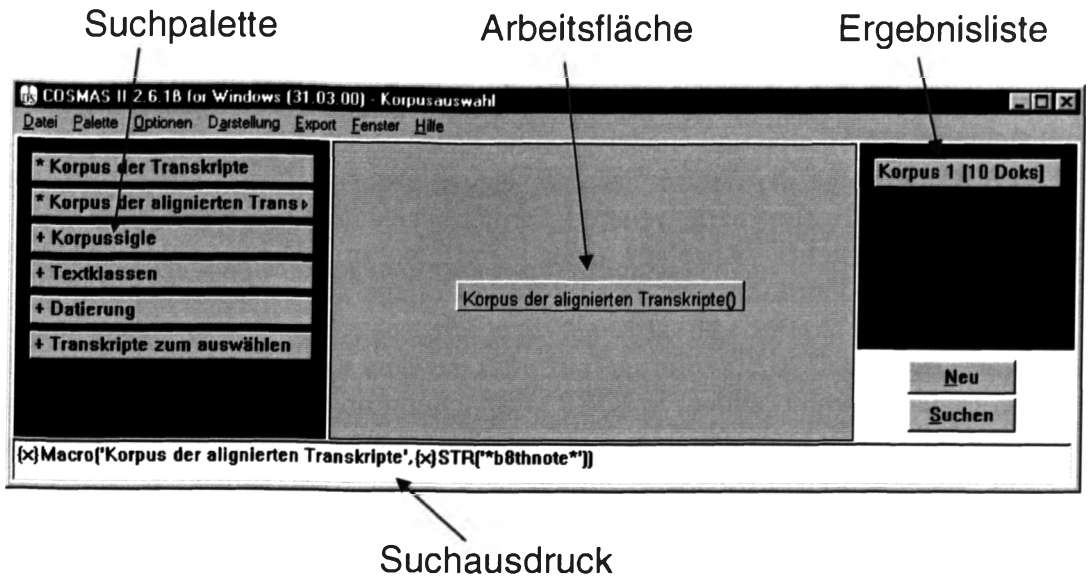


Abb.: Bildschirmfoto des COSMAS-II-Client –Korpusauswahl

- **Suche**: Als nächster Schritt wird im ausgewählten Korpus entweder mit voreingestellten Makros (z.B. „Wort" oder „Äußerung-Pers-Geschlecht") oder mit einer selbstformulierten Suchanfrage gesucht.

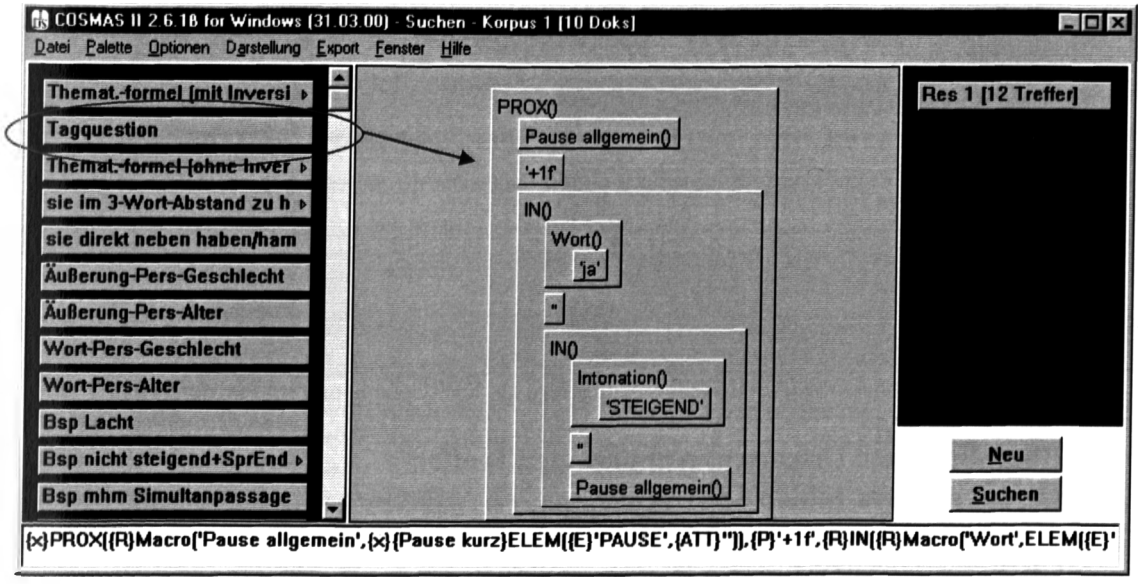


Abb.: Bildschirmfoto des COSMAS-II-Client – Suchanfrage; das vordefinierte Makro „Tag question" wird auf der Arbeitsfläche grafisch und als Suchausdruck in Klammerschreibweise dargestellt.

Der COSMAS-II-Bildschirm ist für die eigentliche Suche ähnlich aufgebaut wie der für die Dokumentauswahl: Man zieht einen Eintrag aus der Suchpalette auf die Arbeitsfläche, startet die Anfrage, wonach das Ergebnis in einer Ergebnisliste erscheint.

Die auf der Arbeitsfläche in Alltagssprache formulierte Suchanfrage wird in einer Extra-Zeile im unteren Teil des COSMAS-II-Fensters auch noch explizit als Suchausdruck in Klammerschreibweise dargestellt. Als Beispiel ist im obigen Bildschirmfoto das Suchmakro „tag question" (d.h. Vergewisserungsfrage) markiert; dabei wird in den Transkript-Äußerungen nach Folgen gesucht, die aus einer Pause, „ja" mit steigender Intonation und wiederum einer Pause bestehen.[12]

- **KWIC**: Anzeige der Treffer in einer einzeiligen Kurzform (Transkriptname und Wortlaut des Treffers ohne weitere Informationen aus dem Transkript).
- **Dokumentansicht:** Anzeige eines Treffers in einem dem DIDA-Editor angenäherten Partiturformat. Die folgende Collage aus zwei Bildschirmfotos zeigt den Recherchetreffer „äh also auf den wir zusteuern ja ich glaube" aus dem Transkript „4050.202,friedfertigefrau", das aus dem Korpus „GF1" (= Gespräche im Fernsehen) stammt, einmal in KWIC-Darstellung und einmal in Partiturdarstellung mit zusätzlichen Informationen (z.B. zum Sprecher, zur interaktiven Einbettung und zur Intonation) und in einem definierbaren größeren Kontext:

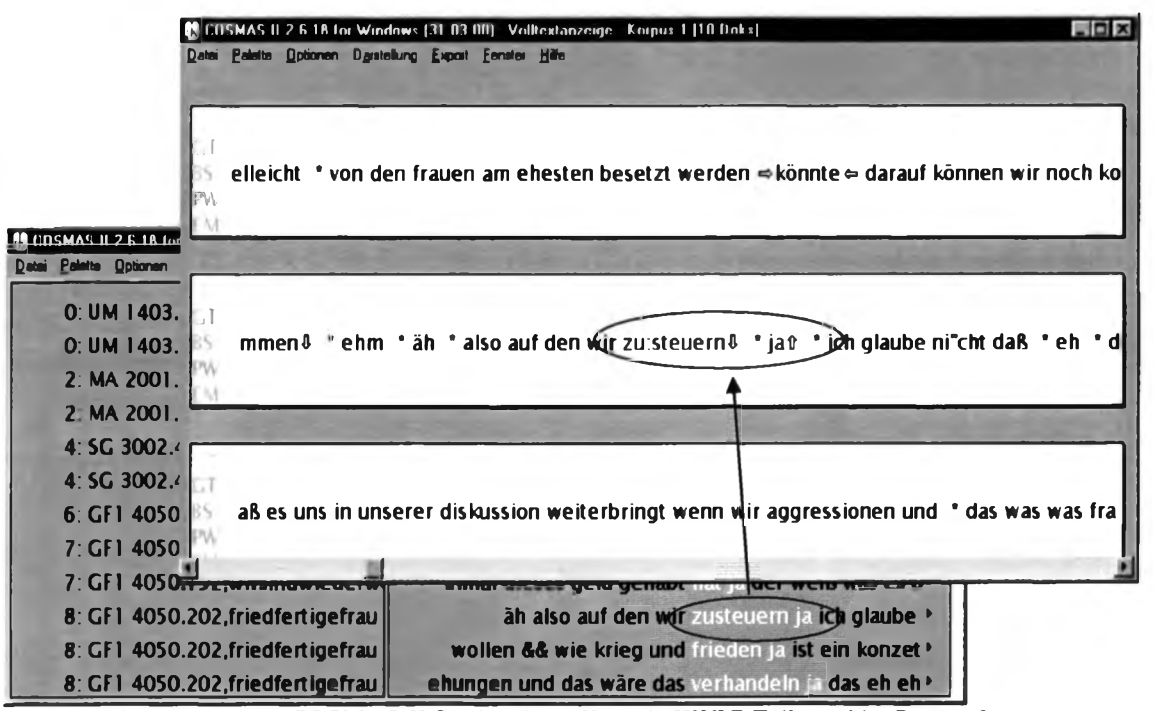


Abb. Anzeige von COSMAS-II-Recherchetreffern als KWIC-Zeile und im Partiturformat

- ggf. (bei alignierten Diskursen) **Anhören der Treffer**;
- **Sortieren und Auswählen** der Treffer;
- **Export** der Treffer in eine Textdatei;
- (ggf. bei alignierten Diskursen) **Weiterverarbeitung der Treffer** im Ausschnitt aus dem Sprachsignal (z.B. Prosodie, Spektrogramm).

[12] Dies ist nur eine ad hoc formulierte einfache COSMAS-II-Anfrage; „tag questions" können natürlich auch mit „ne?" „nicht?" (oder „nich?") realisiert werden und werden nicht notwendig durch Pausen abgetrennt. Die Suchanfrage könnte also durch alternative Realisierungsformen angereichert werden.

In COSMAS II hat man bei DIDA-Transkripten u.a. systematisch folgende Suchmöglichkeiten:

- einzelne Wörter,
- Position von Wörtern in Äußerungen,
- Wörter und andere Zeichenketten in definiertem Abstand zueinander,
- prosodische Notationen und Pausen,
- beliebige Kombinationen dieser Suchelemente.

Hier noch zwei Beispiele für Ergebnisse einer gesprächsanalytisch motivierten COSMAS-II-Suche, die das Ausgabeformat verdeutlichen sollen:

1. „**Tag questions**" (= Vergewisserungsfragen, mit denen der Sprecher eine Bestätigung oder Rezeptionssignale des oder der Adressaten einfordert, mit denen also die etablierte Rederechtsverteilung „einer redet, die anderen hören zu" ratifiziert werden soll); die Suche ergibt im kleinen Korpus der alignierten Transkripte 14 Treffer, hier als KWIC-Treffer exportiert:

```
No. of Lines in KWIC = 14
Query is = 'ePAUSE  /W1 xLEMMA=ja /?0 eSHIFT & xHERE=STEIGEND /?0
ePAUSE '
No. of characters to the left  = 15
No. of characters to the right = 40
============================================================
------------------------------------------------
     21: UMTranscript: 1403.01,erziehung &b8thnotes;
------------------------------------------------
ch dadurch das tempo ja das psychische tempo beim arbeiten in d
h der muss das doch ja äh das is also mit sicherheit ne ne ne
------------------------------------------------
     52: MATranscript: 2001.26,zehner &b8thnotes;
------------------------------------------------
eiße haus hmhm net ja un die wert was is mi=m neckarhafe und
 ohre gewesche net ja un zwar de des find isch besonders perf
------------------------------------------------
    195: SGTranscript: 3002.46,ueblenachr &b8thnotes;
------------------------------------------------
überhaupt kein diskussionspunkt ja und wenn wir hier einen
vergleich schli
ung ihrerseits entgejenzunehm ja äh ferner bin ich bereit als
schiedsman
------------------------------------------------
    261: GF1Transcript: 4050.026,abtreibung &b8thnotes;
------------------------------------------------
das an ja sehr gern ja ja freu ich mich sehr wie is das denn m
n misstand zur krankheit ja und wenn sie jetzt mal weiterdenken
nei
------------------------------------------------
    268: GF1Transcript: 4050.058,demontage &b8thnotes;
------------------------------------------------
 rer wir haben hier ja zu dem stichwort is richtig schreiben d
------------------------------------------------
    275: GF1Transcript: 4050.192,wirsindwiederwer &b8thnotes;
------------------------------------------------
n seit vierzig jahren ja und die mehrheit des volkes hat jeweils
es geld gehabt hat ja der weiß was es bedeutet marktwirtschaf
------------------------------------------------
```

```
    278: GF1Transcript: 4050.202,friedfertigefrau &b8thnotes;
-------------------------------------------------------------
so auf den wir zusteuern ja ich glaube nicht daß eh daß es uns in
u
 wie krieg und frieden ja ist ein konzept von frieden friedfertig
d das wäre das verhandeln ja das eh eh rechtspositionen
einpflanzen
```

Diese Vorkommen von Vergewisserungsfragen können nun zwecks Vergleich und Sprecheridentifizierung an ein Programm zur Prosodieanalyse wie „Praat" übergeben und dort z.B. mit Grundfrequenzanalysen und Spektrogrammen näher bestimmt werden.

2. **„Thematisierungsformeln"** (mit zugelassener Inversion):"es ... geht ... um" (mit zugelassenen Wortabständen von maximal drei Wörtern, um syntaktisch angereicherte Konstruktionen, z.B. „es geht jetzt/mir/vor allem um..." mit zu erfassen. Derartige Formeln ergeben in einem ersten schnellen Zugriff Hinweise auf das Gesprächsthema und auf Gesprächsphasen, in denen das Thema Gegenstand einer Aushandlung zwischen den Gesprächsbeteiligten, möglicherweise auch strittig, ist. Typischerweise tauchen solche Formeln in kontroversen, argumentativen Gespräch im weiteren Verlauf auf, wenn aus Sicht eines der Beteiligten das Gespräch in eine falsche, nicht mit den vordefinierten oder eingangs ausgehandelten Themen kompatible Richtung zu laufen droht – mit „es geht um..." versuchen Beteiligte dann, die verabredete thematische Ausrichtung des Gesprächs einzuklagen; sie benutzen dann oft kontrastierende Formelpaare wie „es geht nicht um..., es geht vielmehr um...". Hier der Anfang einer exportierten Liste von KWIC-Belegen zu Thematisierungsformeln aus den IDS-Gesprächskorpora; dargestellt sind Auszüge aus Beratungsgesprächen:

```
No. of Lines in KWIC = 134
No. of exported lines limited by user to 100
Query is = 'wes /P3 wgeht /P3 OR_OBJECT'
No. of characters to the left  = 15
No. of characters to the right = 40
============================================================
-----------------------------------------
      3: UMTranscript: 1400.06,saussureref
-----------------------------------------
nt hat sondern es geht einfach um die wahrheit die in dem text is
und die
-----------------------------------------
     15: UMTranscript: 1401.03,mieterhoehung
-----------------------------------------
rau dachgarten es geht hier ja um an und für sich um keinen äh
weltbewege
-----------------------------------------
     29: UMTranscript: 1405.01,raetselh-krankh
-----------------------------------------
      nein mhm es geht um was ganz anderes wir werden davon ausge
-----------------------------------------
     33: UMTranscript: 1405.06,schwangernach19
-----------------------------------------
 jetzt net daß es speziell um die krankheit geht ja des wußt ich
nicht das geht ganz gen
```

```
-----------------------------------------
     36: UMTranscript: 1406.06,wiedereingliedg
-----------------------------------------
aber ich meine es geht ja um sie und sie sie möchten ja eben
wieder
-----------------------------------------
     41: UMTranscript: 1408.03,familiengrab
-----------------------------------------
  nein nein äh es geht auch um die aufregung ja auch stimmt haben
sie
-----------------------------------------
     43: UMTranscript: 1409.19,enkelkind
-----------------------------------------
nd zwar is dat es geht um den mike michael da is ja vaterschaft i
-----------------------------------------
     55: unbekanntTranscript: 2003.25,sander
-----------------------------------------
 klar also nee es geht nur so um nen allgemeinen eindruck so um
die atmo
[...]
```

6 Ausblick

Ziel der Datenbankrecherche in Gesprächskorpora ist in der Regel die Mustererkennung: Welche sprachlichen Muster (also etwa Formulierungen, prosodische Merkmale bestimmter Äußerungssegmente sowie deren Kookkurrenzen) sind universell, welche korrelieren mit bestimmten Interaktionstypen, mit anderen Merkmalen, die für bestimmte Teilkorpora von Gesprächen konstitutiv sind, oder mit soziodemografischen Sprechermerkmalen?
Aufnahmen aus authentischen natürlichen Gesprächssituationen sind für die gesprächsanalytische und pragmatische Forschung eine unverzichtbare empirische Grundlage. Transkripte stellen dabei notwendige Hilfsmittel dar. Ihre Zuverlässigkeit und Genauigkeit ist vom Aufwand bei der Erfassung und Korrektur sowie von wechselnden Forschungsinteressen abhängig. Transkripte sollten deshalb immer nur im Zusammenspiel mit den Aufnahmen genutzt werden, an denen die Beschreibungen und Analysen letztlich überprüft werden müssen. In der Forschungspraxis war freilich bislang der Rückgriff vom Transkript auf bestimmte Stellen in der Aufnahme umständlich und unterblieb darum oft – mit dem Resultat transkriptinduzierter Ungenauigkeiten oder Fehldeutungen. Über einen gezielten Zugriff auf Ausschnitte aus dem Sprachsignal mit der Möglichkeit, diese Ausschnitte wiederholt anzuhören und computergestützt weiter zu analysieren, wird diese Ungleichgewichtigkeit aufgehoben: Transkript und Aufnahme kommen bei der Analyse gleichermaßen zu ihrem Recht. Eine Recherche mit COSMAS II ermöglicht zudem einen Zugriff auf eine Vielzahl von gleichartigen Belegen aus dem Gesamt-Korpus oder aus beliebig zusammengestellten Teilkorpora.
In diesem Sinne haben Sprachtechnologie und Gesprächsanalysen gemeinsame Interessen; für Gesprächsforscher können sehr große Korpora natürlicher und dialogischer Gespräche, die sprachtechnologisch aufbereitet werden können, für neue Anstrengungen motivierend sein, Korpora zusammenzustellen und sie in einem Datenbankformat zu verwalten. Indem sie an der Entwicklung von Korpustechnologie und der Modellierung dialogischer Sprechsprache zum Zwecke statistisch basierter

automatischer Analyseverfahren teilnehmen, verschaffen sich Linguisten einen Zugang zu aktuellen Entwicklungsmöglichkeiten für das Retrieval und die Analyse von Texten.[13]

Für die Kriminalistik ergibt sich die Möglichkeit, in einem aktuellen Tondokument vorgefundene Muster, etwa auf der Formulierungs- oder der Prosodie-Ebene, mit Belegen aus einem für den jeweiligen Untersuchungszweck speziell zusammengestellten Korpus zu vergleichen. So lassen sich Ähnlichkeiten, Idiosynkrasien und Unterschiede feststellen, und so können auch Hypothesen, die zu einer Identifizierung und soziodemografischen Zuordnung von Sprechern führen sollen, verifiziert oder falsifiziert werden.

Literatur

Becker-Mrotzek, M. & C. Meier (1999): „Arbeitsweisen und Standardverfahren der Angewandten Diskursforschung". In: Brünner, G., R. Fiehler & W. Kindt, Angewandte Diskursforschung. Band 1: Grundlagen und Beispielanalysen. Opladen/Wiesbaden, 18-45.

Brünner, G., R. Fiehler & H. Wiegers (1996): „Gesprächsmitschnitte 'Geiselnahme Schliersee': Beobachtungen und Fragestellungen für die Analyse". In: Polizei-Führungsakademie, Hrsg., Schlußbericht über das Seminar 'Polizeiliche Kommunikation in extremen Einsatzlagen, 91-94.

Deppermann, A. (1999): Gespräche analysieren. Opladen.

Kallmeyer, W. (1997): „Vom Nutzen des technologischen Wandels in der Sprachwissenschaft: Gesprächsanalyse und automatische Sprachverarbeitung". Zeitschrift für Literaturwissenschaft und Linguistik 107, 124-149.

Rapp, S. (1995): „Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov Models. An Aligner for German". Proceedings of ELSNET goes east and IMACS Workshop „Integration of Language and Speech in Academica and Industry", Moscow. Im Internet unter http://www.ims.uni-stuttgart.de/ ~rapp/aligner.ps.gz.

Selting, M., P. Auer, B. Barden, J. Bergmann, E. Couper-Kuhlen, S. Günthner, U. Quasthoff, C. Meier, P. Schlobinski & P. Uhmann: „Gesprächsanalytisches Transkriptionssystem (GAT)". Linguistische Berichte 173, 91-122.

[13] vgl. Kallmeyer 1997.