

# Grundzüge von EXMARaLDA – einem System zur computergestützten Erstellung und Auswertung von Korpora gesprochener Sprache

Thomas Schmidt

## 1. Einleitung

Dieser Aufsatz gibt einen Überblick über die Grundzüge von EXMARaLDA (EXtensible MARKup Language for Discourse Annotation), einem System zur computergestützten Erstellung und Auswertung von Korpora gesprochener Sprache. Dem Titel dieses Buches folgend, ist sein Anliegen, die wesentlichen Eigenschaften des Systems aus einer diskursanalytischen Perspektive zu motivieren und zu erläutern. Für eine ausführlichere Darstellung der praktischen Möglichkeiten von EXMARaLDA, insbesondere der zugehörigen Softwarewerkzeuge, sei auf Schmidt/Wörner (2005) verwiesen. Eine konkrete Anleitung zur Anwendung von EXMARaLDA für das Transkribieren nach HIAT findet sich in Rehbein et al. (2004). Eine umfassendere Diskussion methodischer und technologischer Aspekte des computergestützten Transkribierens beinhaltet Schmidt (2005).

## 2. Problemstellung

Diskursanalytische Untersuchungen gesprochener Sprache basieren in aller Regel auf Transkriptionen, wobei die Methoden des Erstellens und Auswertens von Transkriptionen aus einer Zeit stammen, in der der Computer als Arbeitsinstrument eines Linguisten noch unüblich war. In seiner ursprünglichen Fassung ist das in der Diskursanalyse gebräuchliche Transkriptionsverfahren – die Halbinterpretativen Arbeitstranskriptionen (HIAT, Ehlich/Rehbein 1976) und ihre Erweiterungen (HIAT 2, Ehlich/Rehbein 1979a,b) – daher auf die Arbeit mit Papier und Bleistift bzw. Schreibmaschine ausgelegt.

Da in HIAT das Verfahren der Partiturnotation angewandt wird und diese Darstellungsform besondere technische Herausforderungen mit sich bringt, zu deren Bewältigung der Computer vielfältige Möglichkeiten bietet (vgl. Ehlich 1992 und Griebhaber 2000), wurden bereits zu einem vergleichsweise frühen Zeitpunkt Lösungen entwickelt, um HIAT-Transkriptionen mit Hilfe des Computers anzufertigen. Für DOS- und später für Windows-Rechner war dies die Software HIAT-DOS (Schneider 2001), für Macintosh-Rechner das Programm syncWriter (Rehbein et al. 1993). Mit dieser Software wurden im Laufe der 90er Jahre zahlreiche größere Diskurskorpora erstellt. Mit der Fortentwicklung der Computertechnologie und damit einhergehenden wachsenden Ansprüchen an die Computerunterstützung in der linguistischen Arbeit offenbarten sich jedoch im Laufe der Zeit einige Mängel und Beschränkungen der mit HIAT-DOS und syncWriter erstellten Korpora.

So machen es die von diesen Programmen verwendeten Datenformate schwierig bis unmöglich, einmal erstellte Daten auch in anderen Umgebungen, etwa auf einem anderen Betriebssystem oder in einer anderen Software, weiter zu verwenden. Ein Austausch von Daten innerhalb der Forschergemeinschaft wird dadurch wesentlich erschwert. Unmittelbar damit verbunden ist die Schwierigkeit, diese Daten auf längere Zeit, etwa für eine dauerhafte Archivierung, nutzbar zu halten, denn dies würde eine Kompatibilität nicht nur mit gegenwärtigen, sondern auch mit zukünftigen Softwareumgebungen voraussetzen.

Da die Schriftverarbeitung auf früheren Computersystemen noch stark auf das Englische fixiert war, ist der Einsatz von HIAT-DOS und syncWriter weiterhin beim Transkribieren solcher Sprachen problematisch, deren Schriftsysteme Erweiterungen des lateinischen Alphabets beinhalten (etwa das Türkische, das Portugiesische oder das Isländische). Für nicht-

lateinische Alphabete (etwa das Kyrillische oder das Griechische, aber auch IPA) oder nicht-alphabetische Schriftsysteme (etwa die CJK-Systeme) potenzieren sich diese Probleme.

Auch die heute als selbstverständlich empfundenen Möglichkeiten, digitale Dokumente zu einem interaktiv navigierbaren Netzwerk zu verknüpfen (Stichwort: Hypertext), sowie digitales Schrift-, Ton- und Bildmaterial in einer Einheit zu integrieren (Stichwort: Multimedia), waren zum Zeitpunkt der Entwicklung von HIAT-DOS und syncWriter kaum gegeben. Beide Programme machen daher nur sehr begrenzt von diesen Möglichkeiten Gebrauch.

Schließlich bieten sich für Korpora, die mit diesen Programmen erstellt wurden, auch nur eingeschränkte Möglichkeiten der computergestützten Auswertung. Software-Werkzeuge, um solche Korpora gezielt nach bestimmten transkribierten Phänomenen zu durchsuchen oder Transkribiertes (teil-)automatisch durch Annotationen zu ergänzen, existieren kaum und sind aufgrund der bereits erwähnten Schwierigkeiten des Datenaustauschs auch nur schwer nachträglich zu entwickeln.

Das hier dargestellte System EXMARaLDA hat zum Ziel, diese Mängel und Beschränkungen zu überwinden. Die Entwicklung des Systems orientiert sich einerseits wesentlich an den bestehenden und bewährten Transkriptionsverfahren der Diskursanalyse. Insbesondere beinhaltet sie die Vorgabe, ältere, mit HIAT-DOS und syncWriter angefertigte Korpora durch geeignete Konvertierungsverfahren wieder nutzbar zu machen und die Erstellung neuer Datenbestände nach dem HIAT-Verfahren zu ermöglichen. Andererseits ordnet sich das System aber auch in zwei größere, über die Diskursanalyse hinausgehende Zusammenhänge ein: Erstens soll es außer für diskursanalytische Daten auch für andere, am Hamburger Sonderforschungsbereich ‚Mehrsprachigkeit‘ vorhandene Korpora gesprochener Sprache anwendbar sein. Dies betrifft insbesondere Daten aus der syntax- oder phonetikorientierten Spracherwerbsforschung, die jeweils nach Transkriptionsverfahren angefertigt werden, die für diese Zwecke optimiert sind. Zweitens ordnet sich die Entwicklung von EXMARaLDA in einen größeren Rahmen aktueller Initiativen ein, die allgemein die Entwicklung von texttechnologischen Verfahren und Standards zur Verarbeitung von Sprachkorpora zum Ziel haben. Exemplarisch seien hier die Text Encoding Initiative (TEI, Johansson 1995), die Language Archiving Technology des MPI in Nijmegen (<http://www.lat-mpi.eu/>) und das Verbundprojekt „Nachhaltigkeit linguistischer Daten“ (Schmidt et al. 2006) genannt.

### 3. Grundprinzipien

Um die genannten Ziele zu erreichen, insbesondere um eine Austauschbarkeit und langfristige Nutzbarhaltung von Korpora gesprochener Sprache zu gewährleisten, folgt die Entwicklung von EXMARaLDA einigen Grundprinzipien.

Als übergeordnetes Prinzip kann dabei die *Datenzentriertheit* gelten. Wie nicht zuletzt die Erfahrungen mit HIAT-DOS und syncWriter gezeigt haben, ist davon auszugehen, dass es langfristig in erster Linie die Daten, und erst in zweiter Linie die zu ihrer Erstellung verwendeten Softwarewerkzeuge sind, die es zu erhalten gilt. Im Zentrum der EXMARaLDA-Systemarchitektur stehen daher Datenmodell und Datenformate, während Eingabe- und Analysesoftware als wenn auch wichtige, so doch prinzipiell austauschbare Werkzeuge zu deren Bearbeitung angesehen werden. Drei weitere Grundprinzipien ergeben sich unmittelbar aus der Datenzentriertheit des Ansatzes:

Erstens wird eine *Unterscheidung* getroffen zwischen *Inhalt und Form* einer Transkription. Diese in der Texttechnologie etablierte Unterscheidung (vgl. Witt 2002) baut auf der Erkenntnis auf, dass die Form eines Dokuments (also die Anordnung und Formatierung von Zeichen auf einem graphischen Medium) durch seinen Inhalt (also die Bedeutung dieser Zeichen) motiviert ist. Im Gegensatz zum Papier erlaubt das digitale Medium, die Trennung dieser beiden Aspekte nicht nur gedanklich zu vollziehen, sondern sie für eine flexiblere Art der Dokumentverarbeitung konkret fruchtbar zu machen. So können bei einer geeigneten digita-

len Kodierung eines Dokuments nach inhaltlichen Gesichtspunkten aus ein und demselben Inhalt mehrere Visualisierungen berechnet werden. Beispielsweise kann die Visualisierung zweier als gleichzeitig kodierter Ereignisse einmal – wie in HIAT – durch eine parallele Anordnung von Zeichenketten in einer Partiturfläche erfolgen und einmal – wie in konversationsanalytischen Transkriptionsverfahren üblich – über eine Verwendung von Klammern in einer zeilenweisen Notationsform.

Zweitens wird von einer *Drei-Ebenen-Architektur* der Datenverarbeitung ausgegangen. Dieses in der Datenbanktheorie gebräuchliche Konzept (vgl. Date 1995) unterscheidet zunächst zwischen einer Daten- und einer Anwendungsebene, wobei die oben erwähnten Software-Werkzeuge und Visualisierungen letzterer zuzurechnen sind. Die Datenebene wird weiter in eine physikalische und eine logische Ebene unterteilt. Erstere betrifft dabei die konkrete Form der physikalischen Speicherung von Daten, letztere abstrakte Eigenschaften von Einheiten und Beziehungen im zugrunde gelegten Datenmodell. Wie weiter unten näher ausgeführt, kann etwa für Transkriptionsdaten auf der logischen Ebene ein Modell in Form eines gerichteten, azyklischen Graphen angenommen werden, das auf der physikalischen Ebene als XML-Datei repräsentiert ist. Indem die logische Ebene in dieser Weise nicht nur über konkrete Anwendungen, sondern auch über das konkrete Speicherungsformat abstrahiert, fördert sie eine zusätzliche Unabhängigkeit der Daten von spezifischen technischen Umgebungen.

Bei der physikalischen Speicherung von Daten wird dann drittens so weit wie möglich von *offenen Standards* Gebrauch gemacht. Hier ist zum einen *Unicode* relevant, ein Standard zur Kodierung von Schriftzeichen, der die allermeisten gebräuchlichen Schriftsysteme, sowie viele weitere für das Transkribieren wichtige Symbole (z.B. frei kombinierbare Diakritika) beinhaltet. Zum anderen liefern die *eXtensible Markup Language (XML)* und mehrere mit ihr verbundene oder auf ihr aufbauende Standards (z.B. XSLT, XPath, XHTML, SVG) die Grundlage zur Kodierung und Verarbeitung strukturierter Dokumente. Sowohl Unicode als auch XML sind mittlerweile in sehr vielen kommerziellen und akademischen Anwendungsbereichen etabliert, so dass ihre Verwendung eine fast ideale Voraussetzung für eine dauerhafte Nutzbarhaltung auch von linguistischen Daten bietet.

Wo es nicht um die Beschreibung und Speicherung der Daten, sondern um die Implementierung geeigneter Software geht, besteht schließlich ein letztes Grundprinzip in der Verwendung *plattformübergreifender Technologien*. Wie wiederum durch die Erfahrungen mit HIAT-DOS und syncWriter (negativ) belegt, ist es eine unabdingbare Voraussetzung für den dauerhaften Erfolg spezialisierter linguistischer Software, dass sie sich in gleicher Weise auf unterschiedlichen Betriebssystemen einsetzen lässt. EXMARaLDA-Software wird daher in JAVA implementiert, einer Technologie, die es dem Entwickler ermöglicht, plattformübergreifend verwendbaren Programmcode zu produzieren. Über eine zwischengeschaltete sog. virtuelle Maschine wird dieser Code auf allen derzeit gängigen Plattformen – Windows, MAC OS, Linux und Unix – lauffähig.

## 4. Datenmodell

### 4.1. Ansatz: Annotationsgraphen

Wie bereits erwähnt, steht im Zentrum der EXMARaLDA-Systemarchitektur ein abstraktes Modell diskursanalytischer Daten. Dieses muss die in einer Transkription enthaltenen Informationen in einer für den Computer eindeutig interpretierbaren und somit algorithmisch manipulierbaren Form repräsentieren. Dazu werden strukturelle Eigenschaften, die allen Transkriptionen gemein sind, identifiziert und in einen algebraischen Formalismus übersetzt. Operationen auf Transkriptionsdaten können dann als Operationen auf dieser Algebra ausgedrückt werden, wodurch in der oben beschriebenen Weise eine weitestgehende Unabhängigkeit von spezifischen technischen Umgebungen erreicht wird.

Als Ausgangspunkt für das EXMARaLDA-Datenmodell dient der von Bird/Lieberman (2001) vorgeschlagene Annotationsgraphenformalismus (AG). Dessen Grundidee beschreiben sie (S. 47) wie folgt:

All annotations of recorded linguistic signals require one unavoidable basic action: to associate a label, or an ordered set of labels, with a stretch of time in the recording(s).

Ausgehend von diesem universellen Zeitbezug aller Transkriptionsdaten besteht der AG in einer Algebra, in der eine Transkription als gerichteter, azyklischer Graph formuliert wird, dessen Knoten zeitlich geordnete Punkte in der Aufnahme repräsentieren und dessen Kanten die zugehörige nicht-zeitliche Information (also etwa die einzelnen Bestandteile des „Transkriptionstexts“) tragen. Wie Bird/Lieberman (2001) selbst betonen, liegt eine Stärke dieses Ansatzes in seiner Einfachheit und der Tatsache dass er „ontologically parsimonious“ (S. 55) ist, also über den objektiv feststellbaren Zeitbezug hinaus keinerlei theoretisch strittigen Annahmen über die in einer Transkription bzw. in transkribierten Diskursen vorkommenden Einheiten macht. Wie in Schmidt (2005) näher ausgeführt, erweist es sich für eine praktische Anwendung allerdings doch als notwendig, den Formalismus einerseits strukturell einzuschränken und andererseits bis zu einem gewissen Grad semantisch zu spezifizieren. Die nächsten Abschnitte beschreiben, wie in EXMARaLDA diese Einschränkung und Spezifizierung vorgenommen wird.

## 4.2. Grundmodell: Basis-Transkription

Zur Illustration des Grund-Datenmodells von EXMARaLDA betrachte man zunächst das in Abbildung 1 wiedergegebene Beispiel einer HIAT-Transkription (vgl. Ehlich 1992):

DS [sup]	<i>faster</i>
DS [v]	Okay. Très bien, très bien.
DS [en]	Okay. Very good, very good.
DS [nv]	<i>right hand raised</i>
FB [v]	Alors ça dépend ((cough)) un petit peu.
FB [en]	That depends, then, a little bit
FB [pho]	[ɛ̃tipø:] ⋮

Abbildung 1: HIAT-Transkription in Partiturnotation

### 4.2.1. Zeitpunkte, Zeitachse, Ereignisse

Nach der Logik des AG lässt sich die Transkription aus Abbildung 1 ausdrücken als ein gerichteter Graph mit 6 Knoten, die jeweils Zeitpunkte im Diskurs bzw. dessen Aufnahme repräsentieren, und 12 diese Knoten verbindenden Kanten, auf denen der Transkriptionstext repräsentiert ist. Abbildung 2 illustriert dies. Dabei wird beispielsweise die Tatsache, dass die Wortfolgen „très bien“ (Sprecher DS) und „Alors ça“ (Sprecher FB) gleichzeitig geäußert werden, dadurch ausgedrückt, dass die entsprechenden Kanten zwischen identischen Start- und Endknoten (2 und 3) gespannt sind. Im EXMARaLDA-Datenmodell heißen die Kanten eines solchen Graphen *Ereignisse*, seine Knoten *Zeitpunkte*. Im Grundmodell, genannt *Basis-Transkription*, wird gefordert, dass zwei beliebige solcher Zeitpunkte immer in eine eindeutige Reihenfolge gebracht werden können; die sich so ergebende geordnete Menge aller Zeitpunkte wird *Zeitachse* genannt.

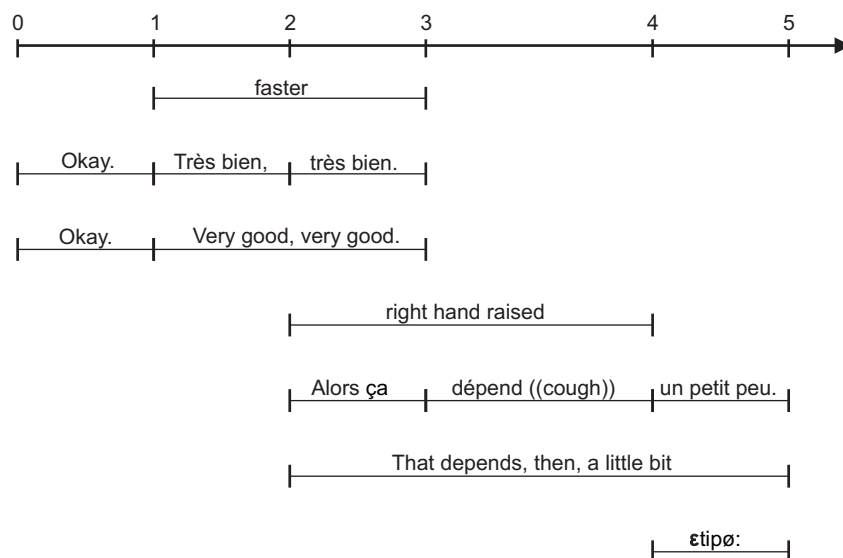


Abbildung 2: Zeitachse und Ereignisse

#### 4.2.2. Spuren: Sprecher- und Kategorienzuordnung, Kategorietypen

Die Ereignisse einer Basis-Transkription können anhand zweier Kriterien geordnet werden: Erstens wird jedes Ereignis einem *Sprecher* zugeordnet. In Abbildung 1 sind diese mit den Kürzeln DS bzw. FB bezeichnet. Zweitens werden Ereignisse anhand ihrer *Beschreibungskategorie* unterschieden. So gibt es für beide Sprecher Ereignisse der Kategorie ‚v(erbal)‘, ‚n(on)v(erbal)‘ und ‚en(glische Übersetzung)‘, für Sprecher FB zusätzlich die Kategorie ‚pho(netische Umschrift)‘. Wie bei der Partiturnotation wird die Sprecher- und -kategorienzuordnung auch im EXMARaLDA-Grundmodell über eine Partition aller Ereignisse in *Spuren* vorgenommen, wobei jede Spur genau einem Paar (Sprecher/Kategorie) entspricht. Innerhalb einer solchen Spur werden dann keine sich überlappenden Ereignisse zugelassen. Die Anzahl der Sprecher ist dabei prinzipiell ebenso unbegrenzt wie die Menge und Bezeichnung der verschiedenen Kategorien. Für die weitere Verarbeitung ist es aber nützlich, drei grundlegend verschiedene *Typen* von Kategorien zu unterscheiden:

- Ereignisse, deren Kategorie zum Typ ‚T(ranskription)‘ gehört, enthalten in der Regel die Beschreibungen der verbalen Sprecherhandlungen. Diese Ereignisse zeichnen sich dadurch aus, dass sie (innerhalb einer Spur) *kombinierbar* und *segmentierbar* sind. Beispielsweise lässt sich das mit „Okay“ beschriebene Ereignis sinnvoll in zwei zeitlich aufeinander folgende Teilereignisse zerlegen, von denen das erste mit „O“, das zweite mit „kay“ beschrieben werden kann. Der umgekehrte Vorgang ist ebenfalls möglich. Diese Eigenschaft, die im Wesentlichen auf dem Prinzip der doppelten Gliederung sprachlicher Zeichen beruht, wird beim Transkribieren z.B. ausgenutzt, um zur Wiedergabe eines „Ins-Wort-Fallens“ die Beschreibung dieses Wortes auf zwei Ereignisse zu verteilen.
- Ereignisse, deren Kategorie zum Typ ‚D(eskription)‘ gehören, enthalten in der Regel die Beschreibungen der nonverbalen Sprecherhandlungen. Im Gegensatz zu Ereignissen des Typs ‚T‘ sind Ereignisse des Typs ‚D‘ nicht kombinierbar oder segmentierbar. Beispielsweise lassen sich aus dem Ereignis mit der Beschreibung „right hand raised“ keine zwei Teilereignisse bilden, die durch Teile der Ausgangsbeschreibung (etwa „right hand“ und „raised“) adäquat beschrieben wären.
- Während Ereignisse des Typs ‚T‘ und ‚D‘ sich jeweils *unmittelbar* auf Phänomene des Gesprächs bzw. von dessen Aufnahme beziehen, stellen Ereignisse, deren Kategorie zum Typ ‚A(nnotation)‘ gehört, zusätzliche Beschreibungen bereits transkribierter Einheiten

dar. Sie besitzen somit nur einen *mittelbaren* Bezug auf die Aufnahme. Im Umkehrschluss bedeutet dies, dass zu jedem Ereignis des Typs ‚A‘ eine Kette von korrespondierenden Ereignissen des Typs ‚T‘ oder ‚D‘ existieren muss, die dem gleichen Start- und Endpunkt auf der Zeitachse zugeordnet ist. Im Beispiel aus Abbildung 2 befindet sich etwa die phonetische Umschrift [ɛ̃tipø:] auf einem Ereignis, dessen Start- und Endpunkt identisch sind mit denen des annotierten Ereignis mit der Beschreibung „un petit peu“.

Abbildung 3 gibt einen Überblick über die Zuordnung der Ereignisse zu Sprechern, Kategorien und Typen.

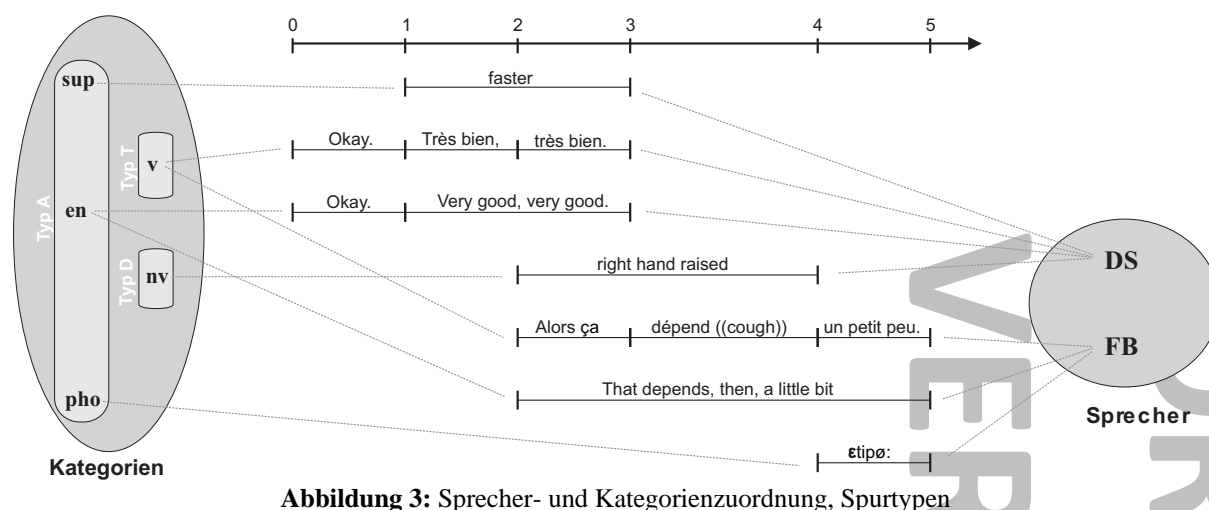


Abbildung 3: Sprecher- und Kategorienz Zuordnung, Spurtypen

#### 4.3. Erweitertes Modell: Segmentierte Transkription

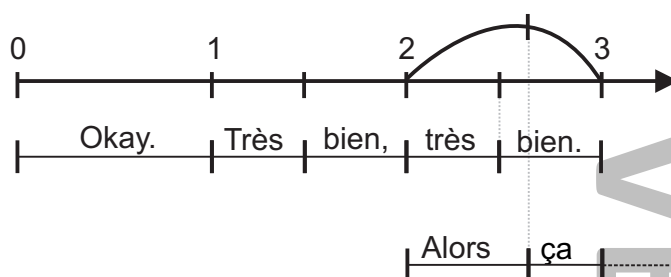
Im Grundmodell richtet sich die Einteilung eines Diskurses in Ereignisse vor allem nach *zeitlichen* Kriterien. Ein Transkribent wählt in der Regel die Einteilung der Zeitachse so, dass bei jeder Änderung der Sprecherkonstellation (z.B. Turnwechsel, Hörersignal) ein entsprechendes Ereignis beginnen kann. Diese zeitliche Unterteilung ist notwendig, um den Diskursverlauf in Form einer Partitur darstellen zu können, und sie ist für eine „manuelle“, d.h. anhand eines Transkripts auf Papier oder auf dem Bildschirm durchgeführte Analyse der Diskursdaten normalerweise auch ausreichend.

Für viele Formen der computergestützten Analyse ist es jedoch wichtig, dass neben den zeitlich motivierten Ereignissen auch *linguistisch* motivierte Einheiten des Diskurses in einer für den Rechner verwertbaren Form *explizit* repräsentiert sind. Für HIAT-Transkriptionen ist in dieser Hinsicht vor allem die Identifikation und Unterscheidung der Einheiten „Wort“, „Äußerung“ und „nicht-phonologische Einheit“ relevant.<sup>1</sup> Beispielsweise kann der verbale Beitrag „Okay. Très bien, très bien.“ von Sprecher DS im obigen Beispiel in zwei Äußerungen und fünf Wörter zerlegt werden, und die Zeichenkette „cough“ im Beitrag von Sprecher FB beschreibt eine nicht-phonologische Einheit. Wie an diesem Beispiel ebenfalls zu ersehen ist, werden die verschiedenen linguistischen Einheiten bereits auf der Zeichenebene implizit voneinander getrennt: Äußerungen sind durch ein Äußerungsendzeichen (Punkt, Fragezeichen

<sup>1</sup> Verschiedene Transkriptionssysteme definieren verschiedene solcher Einheiten. Während die Einheit „Wort“ in allen gebräuchlichen Systemen und in vergleichbarer Weise verwendet wird, variieren Definition und Bezeichnungen der anderen Einheiten von System zu System. Der Einheit „Äußerung“ in HIAT liegen die Überlegungen aus Rehbein (1995) zum Segmentieren zugrunde. Der Begriff „nicht-phonologische Einheit“ wird bereits in Ehlich/Rehbein (1976) verwendet.

etc.) abgeschlossen, Wörter stehen zwischen Leer- bzw. satzinternen Interpunktionszeichen, und nicht-phonologische Einheiten sind durch doppelte runde Klammern gekennzeichnet. Diese in den Transkriptionskonventionen formulierten Regelmäßigkeiten lassen sich ausnutzen, um aus der zeitlich motivierten Ereignisstruktur, die im Grundmodell repräsentiert ist, eine linguistisch motivierte Segmentstruktur automatisch zu berechnen.

Dabei kann die Grundstruktur des Modells beibehalten werden – die zusätzliche Segmentierung wird über das Einfügen zusätzlicher Punkte auf der Zeitachse repräsentiert. Allerdings können aufgrund der gängigen Praxis, bei überlappenden Sprecherbeiträgen nur Anfangs- und Endpunkt der Überschneidung festzuhalten, Paare von Zeitpunkten entstehen, die sich nicht mehr eindeutig zueinander anordnen lassen. Im obigen Beispiel hat der Transkribent zum Beispiel keine zeitliche Ordnung für den Beginn der Wörter „bien“ und „ça“ festgelegt. Um auch solche Fälle noch innerhalb des zeitachsenbasierten Paradigmas ausdrücken zu können, lässt das erweiterte EXMARaLDA-Datenmodell daher eine verzweigende Zeitachse zu.



**Abbildung 4:** Segmentierung von Sprecherbeiträgen in Wörter durch Einfügen zusätzlicher Punkte auf der Zeitachse. Die Zeitachse verzweigt zwischen den Punkten 2 und 3, um die nicht definierte Ordnung der Wörter „bien“ und „ça“ repräsentieren zu können.

Basierend auf einer solchen Segmentierung der Transkriptionsdaten in linguistisch motivierte Einheiten ergeben sich dann weit reichende Möglichkeiten, Transkriptionen computergestützt auszuwerten.

## 5. Umsetzung

Die Umsetzung des EXMARaLDA-Datenmodells besteht in der Definition geeigneter Datenformate, sowie in der Entwicklung von Softwarewerkzeugen zum Eingeben und Bearbeiten von Transkriptionen und zum Erstellen, Verwalten und Auswerten von Transkriptionskorpora. Diese Formate und Werkzeuge werden im Folgenden kurz skizziert.

### 5.1. Datenformate

EXMARaLDA arbeitet mit zwei aufeinander aufbauenden XML-basierten Datenformaten, deren Syntax ist jeweils über eine formale Dokumentgrammatik, eine so genannte Document Type Definition (DTD), spezifiziert ist. Das zum Grundmodell gehörige Datenformat ist die „Basic-Transcription“, in der die oben beschriebene Repräsentation der Transkription als ein System aus Zeitachse, Spuren und Ereignissen festgehalten wird. Dies geschieht in Form von XML-Elementen, die über eine hierarchische Anordnung und Attribute aufeinander bezogen sind. Abbildung 5 illustriert dies.

```

<basic-transcription>
  <head>
    <speakertable>
      <speaker id="SPK0" abbreviation="DS"/>
      <speaker id="SPK1" abbreviation="FB"/>
    </speakertable>
  </head>
  <body>
    <common-timeline>
      <tli id="T0"/>
      <tli id="T1"/>
      <tli id="T2"/>
      <tli id="T3"/>
      <tli id="T4"/>
      <tli id="T5"/>
    </common-timeline>
    [...]
    <tier id="TIE2" speaker="SPK0" category="v" type="t">
      <event start="T0" end="T1">Okay. </event>
      <event start="T1" end="T2">Très bien, </event>
      <event start="T2" end="T3">très bien. </event>
    </tier>
    <tier id="TIE3" speaker="SPK0" category="en" type="a">
      <event start="T0" end="T1">Okay. </event>
      <event start="T1" end="T3">Very good, very good.</event>
    </tier>
    <tier id="TIE4" speaker="SPK0" category="nv" type="d">
      <event start="T2" end="T4">right hand raised</event>
    </tier>
    <tier id="TIE5" speaker="SPK1" category="v" type="t">
      <event start="T2" end="T3">Alors ça </event>
      <event start="T3" end="T4">dépend ((cough)) </event>
      <event start="T4" end="T5">un petit peu. </event>
    </tier>
    [...]
  </body>
</basic-transcription>

```

**Abbildung 5:** Repräsentation des Beispiels in einer „Basic-Transcription“-XML-Datei

In ähnlicher Weise werden im zweiten EXMARaLDA-Datenformat, der „Segmented-Transcription“, die zum erweiterten Modell gehörigen Informationen bzgl. der Segmentierung in linguistische Einheiten festgehalten.

## 5.2. Visualisierungsmethoden

Ungeachtet aller „modernen“ Methoden der voll- oder teilautomatischen, quantifizierenden oder gar statistisch operierenden Analyse von Diskurskorpora, bleibt der Schritt des qualitativen, „manuellen“ Studiums von Transkripten der zentrale Schritt in einer diskursanalytischen Untersuchung. EXMARaLDA unterstützt diese bewährte Arbeitsweise daher durch mehrere Methoden zur Ausgabe von Transkriptionen auf Papier oder Bildschirm.

Am technisch anspruchsvollsten ist eine Ausgabe als Partitur, insbesondere wenn wie in Abbildung 6 die Partiturfläche auf eine bestimmte Seiten- oder Bildschirmbreite umgebrochen werden muss. Der Partitur-Editor (s.u.) nimmt diesen Umbruch automatisch vor, wobei sich Parameter wie Umbruchbreite, Partiturrahmen etc. vom Benutzer bestimmen lassen. Das Ergebnis kann entweder direkt auf einen Drucker ausgegeben werden, oder in Form einer HTML- einer RTF- oder einer SVG-Datei für eine webfähige Veröffentlichung, eine weitere Bearbeitung oder zur Integration in andere Dokumente gespeichert werden.

[11]

GN	Das <u>ä</u> rgert mich ein wenig.
GD	• • • Wir wollen • gleich noch weiter GD: im Folgenden GN zugewandt

[12]

GD	darüber sprechen. Ich hör gerade: Rudi Völler <u>is</u> bei Waldemar GD: im Folgenden der Kamera zu gewandt
----	--

Abbildung 6: Visualisierung einer Transkription als Partitur

Die in Abbildung 7 dargestellte Äußerungsliste ist eine weitere Visualisierungsmethode, die ausgehend von der Segmentierung der Transkription in linguistische Einheiten (hier: Äußerungen) berechnet werden kann:

<b>DS:</b>	Okay.
<b>DS:</b>	Très bien, très bien.
<b>FB:</b>	Alors ça dépend ((cough)) un petit peu.

Abbildung 7: Visualisierung einer Transkription als Äußerungsliste

Diese beiden Ausgabeformen sind geeignet, um sowohl am Computer-Bildschirm als auch auf einem Papierausdruck angezeigt und analysiert zu werden. Beschränkt man sich auf den Bildschirm als Ausgabemedium, erschließen sich weitere Darstellungsformen, die insbesondere von den Hypertext-Fähigkeiten des Rechners Gebrauch machen und so neuartige Formen der Transkriptionsanalyse ermöglichen. Abbildung 8 zeigt, wie eine alphabetische Wortliste, die wiederum ausgehend von der Segmentierung der Transkription in linguistische Einheiten (hier: Wörter) berechnet wurde, mit der Transkription verknüpft werden kann: Ein Klick auf ein Wort in der Wortliste im rechten Bildschirmbereich rollt dabei die Partitur-Darstellung der Transkription in der linken Bildschirmhälfte an die betreffende Stelle. Eine solche Darstellung kann beispielsweise für die Analyse von Wortverwendungen in ihrem jeweiligen Gesprächskontext verwendet werden.

[9]

KLA [v]	warum lacht der <u>denn</u> jetzt?	Naja, is egal.	Ja, sicher.
KLA [en]	laughing?	Doesn't matter.	Yes, sure.
KLA [fr]	ce qu'il rit?	C'est égal.	Oui, bien sûr.
ERW [v]		Hmhm. Ähm, äh jaa. Ach, ach, Volkshochschule?	
ERW [en]		Yes. Ah, yes. Ah, adult education centre?	
ERW [fr]		Oui. Ah, oui. Ah, université de peuple?	
[aud]			

[10]

KLA [v]	Hab ich doch grade gesacht. So n Kursus an ner Volkshochschule. Was? Zimmer vierzehn.		
KLA [en]	That's what I just said. A course at the adult education centre. What? Room fourteen.		
KLA [fr]	C'est ce que j'ai dit. Un cours à l'université de peuple. Quoi? Salle quatorze.		
ERW [v]	Ach, Volkshochschule! Ja, welches Zimmer?		Nee, ich hab
ERW [en]	Ah, adult education centre! Yes, which room?		No, I asked
ERW [fr]	Ah, université de peuple! Oui, quelle salle?		Non, j'ai
[aud]	AUDIO7		

[11]

KLA [v]	Ja, Zimmer vierzehn sach ich!	Ja, Zimmer vier
ET A [en]	Yes, I said room fourteen.	Yes, room

decke	T28
denn	T39 T59 T62 T70 T10
der	T2 T39 T65 T78
des	T0 T94
die	T16 T25 T28
doch	T45 T63 T68 T72
doof	T91
dreizehn	T0 T94
drin	T56
du	T10 T96
egal	T31 T42
einer	T3 T4
einsam	T92
en	T34
erwin	T11 T74
familie	T69
fühlen	T92
ganz	T31

Abbildung 8: Hypertextvisualisierung einer Transkription: Partitur und verknüpfte alphabetische Wortliste

## 5.3. Softwarewerkzeuge

### 5.3.1. Partitur-Editor

Das wichtigste Instrument zum Erstellen von Diskurstranskriptionen ist der Partitur-Editor. Dessen graphische Oberfläche stellt eine Transkription als potentiell unbegrenzt breite Partitur dar. Dabei finden sich die oben beschriebenen Einheiten „Spur“, „Ereignis“ und „Zeitachse“ aus dem EXMARaLDA-Grundmodell in Elementen der graphischen Benutzerschnittstelle sowie in Bedienungselementen des Editors wieder:

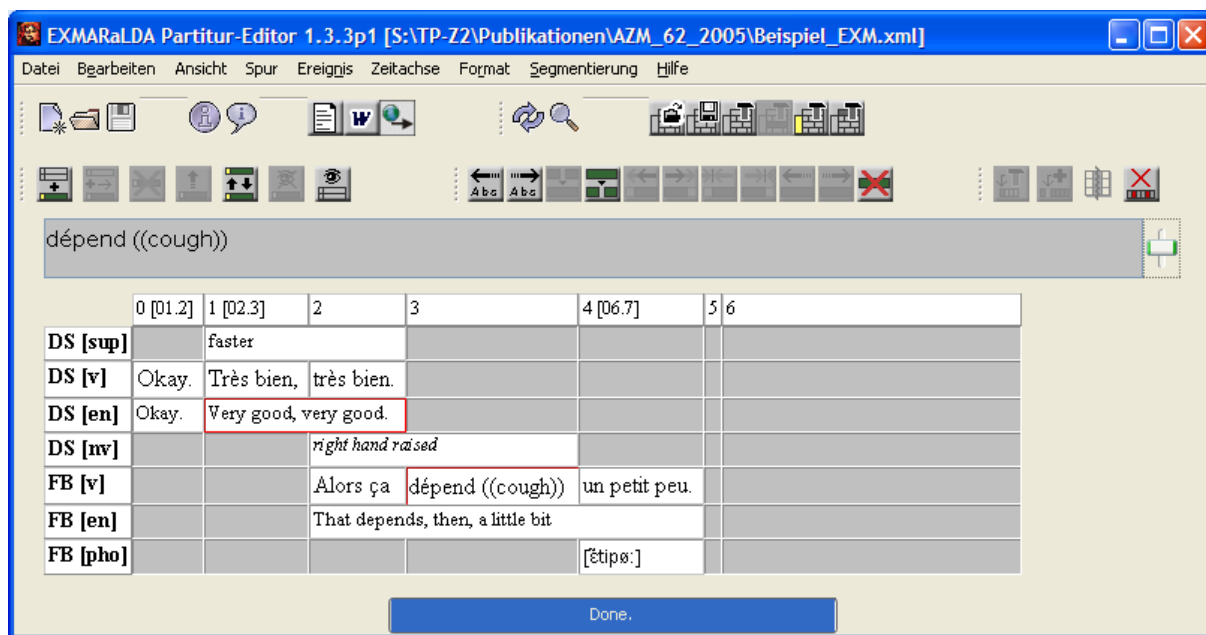


Abbildung 9: Graphische Benutzerschnittstelle des EXMARaLDA Partitur-Editors

Durch die Zuordnung absoluter Zeitwerte zu einzelnen Punkten der Zeitachse kann die Transkription im Editor optional mit einer zugehörigen digitalen Audio- oder Videoaufnahme verknüpft werden. Dies ist einerseits beim Transkribieren selbst hilfreich, weil über diese Synchronisation dem Transkribenden die Navigation in der Aufnahme erleichtert wird. Andererseits eröffnet es auch beim Auswerten die Möglichkeit, von jeder Stelle in der Transkription auf die entsprechende Stelle in der Aufnahme zurückzugreifen, etwa um sich in Zweifelsfällen über die Korrektheit der Transkription rückzuversichern.

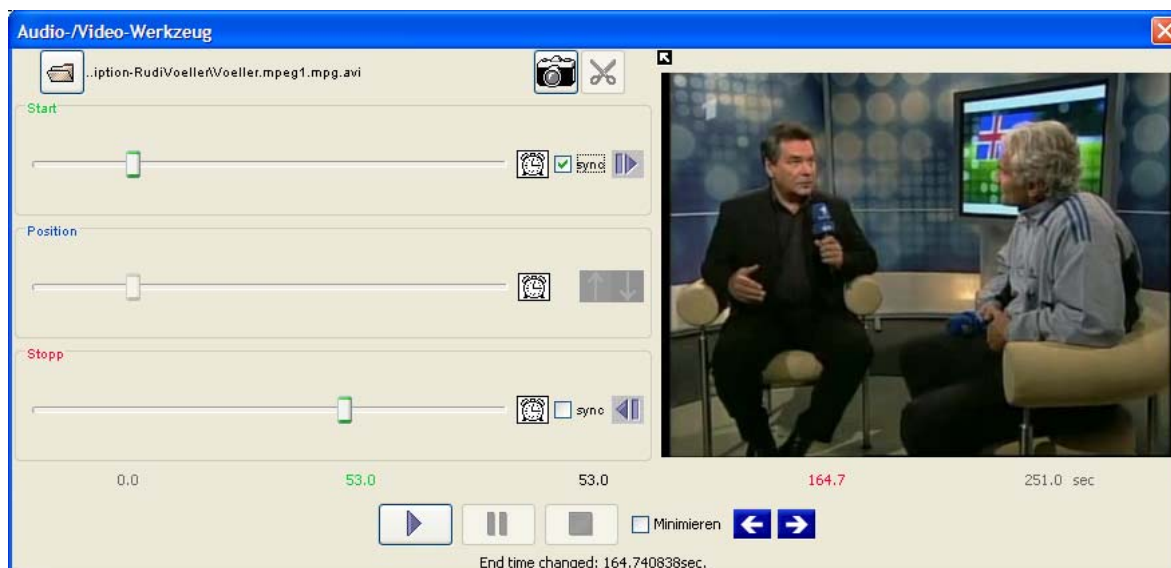


Abbildung 10: Audio-/Video-Werkzeug des EXMARaLDA Partitur-Editors

### 5.3.2. Corpus-Manager

Der Corpus-Manager ist ein Werkzeug, mit dem sich Transkriptionen zu Korpora bündeln lassen. Dazu wird eine – ebenfalls XML-basierte – Korpusdatei angelegt, in der Meta-Informationen zu Diskursen und beteiligten Personen mit den zugehörigen Transkriptions- und Aufnahmedateien verknüpft werden können. Das zugrunde liegende Datenmodell geht dabei davon aus, dass – wie in diskursanalytischen Korpora üblich – Personen in verschiedenen Konstellationen an mehreren Diskursen beteiligt sein können, und dass zu einem Diskurs mehrere Transkriptionen existieren können.

Die Meta-Informationen umfassen biographische Eigenschaften von Sprechern, wie etwa Alter, Geschlecht oder Herkunft und Informationen zu Diskursen, wie etwa Ort oder Diskursart. Indem solche Parameter mit den Transkriptionen verbunden werden, können bei der Auswertung außersprachliche Faktoren systematisch mit sprachlichen Phänomenen in Bezug gesetzt werden.

Der Corpus-Manager unterstützt solche Auswertungen, indem er es dem Benutzer erlaubt, durch geeignete Suchanfragen oder Filter aus einem Gesamtkorpus Teilkorpora zu erstellen, die dann z.B. „alle Diskurse mit bilingualen, weiblichen Sprechern“ oder „alle Diskurse, die vor 1990 in Hamburg aufgenommen wurden“ enthalten.

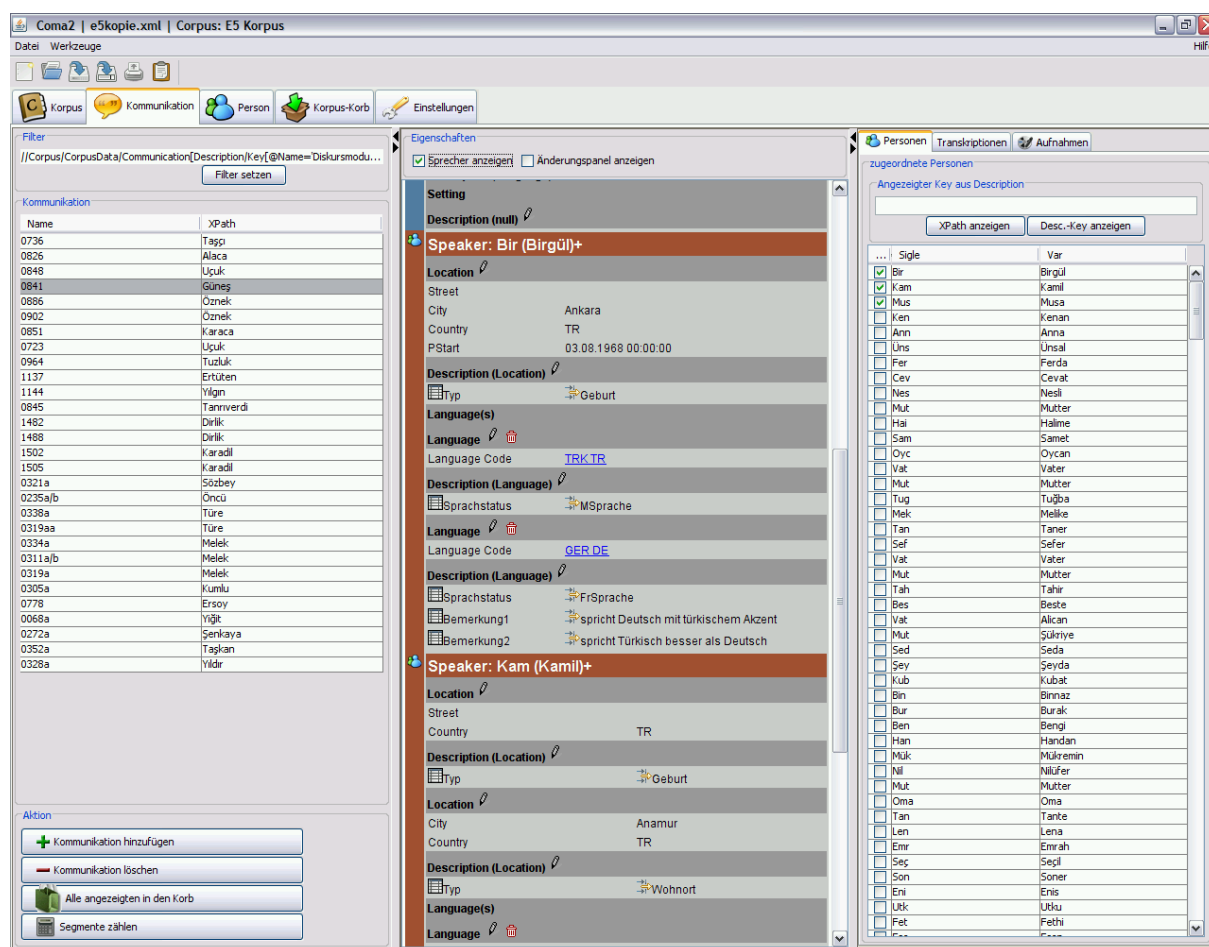


Abbildung 11: Benutzerschnittstelle des Corpus-Manager

### 5.3.3. Suchwerkzeug

Eine sicherlich wesentliche Motivation für den Einsatz des Computers in der Diskursanalyse ist die Möglichkeit, die klassische Einzelfallanalyse von Transkriptionen durch Analysen großer (möglicherweise hinsichtlich ausgewählter Parameter repräsentativer) Korpora mit quantitativen (möglicherweise statistisch evaluierenden) Belegen zu untermauern oder zu ergänzen.

Ein Vorbild für solche Analysen findet sich im Gebiet der so genannten „Korpuslinguistik“, die sich allerdings überwiegend mit geschriebener Sprache auseinandersetzt. Eine für Untersuchungen auf diesem Gebiet typisches Instrument ist ein Konkordanzwerkzeug, mit dessen Hilfe größere Korpora nach ausgewählten Phänomenen durchsucht und die Suchergebnisse in ihrem jeweiligen Kontext ausgewertet (z.B. gezählt, aber auch qualitativ charakterisiert) werden können. Das EXMARaLDA-Suchwerkzeug EXAKT (EXMARaLDA Analyse- und Konkordanz-Tool) überträgt die Arbeitsweise solcher Konkordanzprogramme auf die Arbeit mit Korpora gesprochener Sprache.

Dabei kann der Benutzer ein gegebenes (mit dem Corpus-Manager erstelltes) Korpus über verschiedenartige Suchausdrücke nach transkribierten Phänomenen durchsuchen lassen. Das Ergebnis der Suche wird dann zunächst als eine KWIC (Keyword in Context)-Konkordanz angezeigt. Anders als bei Daten geschriebener Sprache ist für Daten gesprochener Sprache mit einer solchen KWIC-Konkordanz aber noch nicht der vollständige, für die Analyse relevante Kontext der Fundstelle angegeben. Einerseits können auch Handlungen anderer Sprecher, die in zeitlicher Nähe der Fundstelle (also kurz vorher, nachher oder simultan) stattfinden, für die Analyse eine Rolle spielen, andererseits mag die Transkription auch Angaben zu non-verbale Handlungen desselben Sprechers enthalten, die in der KWIC-Konkordanz nicht

sichtbar sind. Aus diesem Grund erlaubt EXAKT dem Benutzer, sich zu jeder Fundstelle deren gesamten transkribierten Kontext in Form einer Partiturdarstellung anzeigen zu lassen. Darüber hinaus kann zu jeder Fundstelle die in den Meta-Informationen des Korpus festgehaltene Information (s.o.) abgerufen werden. Wie oben beschrieben werden so die Ergebnisse auch zu ihrem außersprachlichen Kontext – beispielsweise dem Kontext des gesamten Gespräches oder biographischen Charakteristika der Sprecher – in Bezug gesetzt.

The screenshot displays the EXAKT software interface. The top window shows a KWIC-Konkordanz (KWIC concordance) for the search term 'çocuk'. The table below is a representation of the data shown in the interface:

Communication	Speaker	Left Context	Match	Right Context	Eigensch...
0938	Ize	Ach söd' Aää' ((3s)) Annen	çocuklarla	kavga ediyö müymüş küçükken?	Interviewerin
0938	Ize	Dayın uslu bi	çocuk	müymüş?	Interviewerin
0639	Bir	ğlum orda beni beklıyor. Ben burda	çocuklarla	oynamak için geldim. Sonra işte s	Interviewerin
0639	Bir	Kral, Kraliçe... Onların da	çocukları	da ne olur? Kız olursa ((1s)) Pr	Interviewerin
0826	Bir	an zaman gelirsın buraya. Oynarsın	çocuklarla	.	Interviewerin
0780	Mut	/ bize sormuştu, köyde n'aptınız	çocukluğu...	diye.	Elternteil
0780	Mut	Baban n'apmış	çocukluğu...	, ben n'apmışım, halan n'apmış. ••	Elternteil
0640	Ann	Hih! Bunlar v'ardı. •	çocuk	gibi küçük dimi? Simdi Pamuk Prens	Interviewerin
0640	Ann	Sakalları beyaz. Bunlar yaşlı. •	Çocuk	değil.	Interviewerin
0691	Ken	ayır... bazı okullarda, mesela benim	çocukları	da okula gidiyorlar. ((1,2s)) Ehh,	Interviewerin
0691	Ken	ber? •• Ama diğer sınıfta da Türk	çocukları	var, deme? ((1s)) Onlar Türkçe der	Interviewerin
0691	Ken	••• Kim bilmiyor Türkçeyi? Türk	çocukları	Türkçe bilmiyor, öyle mi?	Interviewerin
0691	Ken	•Anlamadı	çocuk	yâ.	Interviewerin
0691	Ken	O	çocuk	nereli, o Tim?	Interviewerin
0731	Nes	ahseteceğiz biliyor musunuz? Annen	çocukken	neler yapmış?	Interviewerin
0731	Nes		çocukken	, hehe. Ana okulunda farklılık, yad	Interviewerin
0731	Nes	Sen gidemedin mi •	Çocuklukta	?	Interviewerin
0731	Nes	(Ge)zilere felan hiç katıldın mı	çocukken	?	Interviewerin
0731	Nes	şirivordu havuz. •Yürüyerek Tabi	çocukken	daha ağır yürüyösün, yavaş yürüyös	Interviewerin

The bottom window shows a Partitur-Darstellung (Partitur representation) of the selected entry. It displays the transcript of the utterance: "Şimdi nelerden bahseteceğiz biliyor musunuz? Annen çocukken neler yapmış?" with a blue highlight under the word "çocukken". The interface also shows a list of speakers (Nes, Nes [de], Nes [nv], Fik) and a time stamp ((1,5s)) (Okay warte mal!).

Abbildung 12: KWIC-Konkordanz und Partitur-Darstellung in EXAKT

## Literatur

- Bird, Steven / Liberman, Mark (2001): *A formal framework for linguistic annotation*. In: *Speech Communication* 33(1,2), 23-60.
- Ehlich, Konrad (1992): *HIAT - a Transcription System for Discourse Data*. In: Edwards, Jane / Lampert, Martin (Hrsg.): *Talking Data – Transcription and Coding in Discourse Research*. Hillsdale: Erlbaum. 123-148.
- Ehlich, Konrad / Rehbein, Jochen (1976): *Halbinterpretative Arbeitstranskriptionen (HIAT)*. In: *Linguistische Berichte* 45, 21-41.
- Ehlich, Konrad / Rehbein, Jochen (1979a): *Zur Notierung nonverbaler Kommunikation für diskursanalytische Zwecke (HIAT2)*. In: Winkler, Peter (Hrsg.): *Methoden der Analyse von Face-To-Face-Situationen*. Stuttgart: Metzler, 302-329.
- Ehlich, Konrad / Rehbein, Jochen (1979b): *Erweiterte halbinterpretative Arbeitstranskriptionen (HIAT2): Intonation*. In: *Linguistische Berichte* 59, 51-75.

- Grießhaber, Wilhelm (2000): *Verfahren und Tendenzen der funktional-pragmatischen Diskursanalyse. Vom Speiserestaurant zum Cybercafé*. In: Ivanyi, Z. (Hrsg.): *Gesprächsforschung: Tendenzen und Perspektiven*. Frankfurt/M.:Lang, 75-95.
- Johansson, Stig (1995): *The approach of the Text Encoding Initiative to the encoding of spoken discourse*. In: Leech, Geoffrey / Myers, Greg / Thomas, Jenny (Hrsg.) *Spoken English on Computer: Transcription, Markup and Application*. Harlow: Longman, 82-98.
- Rehbein, Jochen / Grießhaber, Wilhelm / Löning, Petra / Hartung, Marion / Bührig, Kristin (1993): *Manual für das computergestützte Transkribieren mit dem Programm syncWRITER nach dem Verfahren der Halbinterpretativen Arbeitstranskriptionen (HIAT)*. Hamburg: Germanisches Seminar, Universität Hamburg.
- Rehbein, Jochen (1995): *Segmentieren*. Verbmobil Memo 64. Hamburg.
- Rehbein, Jochen / Schmidt, Thomas / Meyer, Bernd / Watzke, Franziska / Herkenrath, Annette (2004): *Handbuch für das computergestützte Transkribieren nach HIAT*. Arbeiten zur Mehrsprachigkeit, Folge B (Nr. 56). Hamburg.
- Schmidt, Thomas (2005): *Computergestützte Transkription – Modellierung und Visualisierung gesprochener Sprache mit texttechnologischen Mitteln*. Frankfurt a.M.: Lang.
- Schmidt, Thomas / Wörner, Kai (2005): *Erstellen und Analysieren von Gesprächskorpora mit EXMARaLDA*. In: *Gesprächsforschung* 6, 171-195.
- Schmidt, Thomas / Chiarcos, Christian / Lehmborg, Timm, / Rehm, Georg / Witt, Andreas / Hinrichs, Erhard (2006): *Avoiding Data Graveyards: From Heterogeneous Data Collected in Multiple Research Projects to Sustainable Linguistic Resources*. Proceedings of the E-MELD workshop 2006, Michigan State University in East Lansing, Michigan.
- Witt, Andreas (2002): *Multiple Informationsstrukturierung mit Auszeichnungssprachen. XML-basierte Methoden und deren Nutzen für die Sprachtechnologie*. Dissertation, Universität Bielefeld.

D  
R  
A  
F  
T  
  
V  
E  
R  
S  
I  
O  
N