

Gesprächstranskription auf dem Computer – das System EXMARaLDA

Thomas Schmidt

Abstract

Der Einsatz des Computers zur Transkription natürlicher Gespräche ist in der Praxis zwar weit verbreitet, die schnelle Weiterentwicklung der Computertechnologie hat aber dazu geführt, dass verschiedene Systeme oft scheinbar zusammenhangslos nebeneinander stehen, ohne dass ihre Gemeinsamkeiten und Unterschiede Gegenstand einer umfassenden theoretischen Betrachtung wären. Der vorliegende Aufsatz will einen Beitrag zu einer solchen theoretischen Betrachtung der Gesprächstranskription auf dem Computer liefern, indem er einerseits einige grundlegende Überlegungen zu diesem Thema anstellt, andererseits einige allgemeine Aspekte der Konzeption und Umsetzung des Systems EXMARaLDA, das am SFB "Mehrsprachigkeit" an der Universität Hamburg entwickelt wird, beschreibt.

Keywords: Transkription, Konversationsanalyse, Diskursanalyse, Gesprächsanalyse, Texttechnologie

English Abstract

Although the use of a computer for the transcription of natural conversation is wide-spread in practice, we observe that – due to the rapid development in computer technology – different systems often stand side-by-side in an apparently unrelated manner without their commonalities and differences being subject to theoretical considerations. The presents paper intends to make a contribution to such a theoretical view of computer transcription of conversation: on the one hand, it contains some basic reflections on the subject, and on the other hand it describes some general aspects of the concept and realisation of the system EXMARaLDA which is being developed at the SFB "Mehrsprachigkeit" at the University of Hamburg.

Keywords: Transcription, conversation analysis, discourse analysis, text technology

- Einleitung
- 1. Warum Gesprächstranskription auf dem Computer?
- 2. Schwächen existierender Systeme zur Gesprächstranskription auf dem Computer
 - 2.1. syncWriter
 - 2.2. GAT
- 3. Prinzipien für die Gesprächstranskription auf dem Computer
 - 3.1. Trennung von Inhalt und Darstellung
 - 3.2. Verwendung von Standards
- 4. Aspekte der Umsetzung
- 5. EXMARaLDA
- 6. Ein Beispiel
- 7. Literatur

Einleitung

Der vorliegende Aufsatz befasst sich mit der Transkription natürlicher Gespräche mit Hilfe des Computers. Er tut dies vor dem Hintergrund des Systems EXMARaLDA, das im Rahmen eines Projektes "Mehrsprachige Datenbank" am Sonderforschungsbereich "Mehrsprachigkeit" an der Universität Hamburg entwickelt wird. Ziel des Aufsatzes ist jedoch weniger, dieses System in seinen Einzelheiten vorzustellen – dies ist an anderer Stelle (Schmidt 2002) geschehen –, sondern vielmehr aufzuzeigen, dass einerseits die Praxis der Gesprächstranskription auf dem Computer innerhalb der linguistischen Forschung einen Stellenwert erlangt hat, der es rechtfertigt, sich auf einer theoretischen Ebene mit ihr auseinanderzusetzen, und dass andererseits die bereits vorhandenen Systeme zur Gesprächstranskription auf dem Computer den praktischen Ansprüchen, die sich aus einer solchen theoretischen Auseinandersetzung ergeben, nur unzureichend genügen. Mit anderen Worten: es soll gezeigt werden, dass und warum der Aufwand, den die Entwicklung eines neuen System mit sich bringt, berechtigt ist, und welche Möglichkeiten sich aus der Entwicklung eines solchen Systems für die Forschungspraxis ergeben können.

Im ersten Abschnitt wird daher zunächst eine Antwort auf die Frage gesucht, warum der Einsatz eines Computers zur Gesprächstranskription überhaupt Auswirkungen darauf haben kann, wie transkribiert und mit Transkriptionen gearbeitet wird bzw. werden könnte.

Der zweite Abschnitt beschreibt dann am Beispiel zweier verbreitet genutzter Systeme, warum mit Recht behauptet werden kann, dass die gängige Praxis der Gesprächstranskription ihr theoretisch vorhandenes Potential nicht ausschöpft. Die Wahl der Beispiele – syncWriter und GAT – erfolgt dabei natürlich nicht ganz zufällig. Während die beiden Systeme nämlich zwar zu weitgehend vergleichbaren theoretischen Zwecken eingesetzt werden, existieren in der Praxis weder geeignete Schnittstellen zwischen ihnen noch zu anderen Anwendungen, die auf der computergestützten Gesprächstranskription basieren. Die Gründe hierfür, sowie Möglichkeiten, diesem Umstand Abhilfe zu schaffen, werden in Abschnitt 3 dargelegt.

Die Abschnitte 4 und 5 beschreiben übersichtsartig, wie die in den vorherigen Abschnitten angestellten Überlegungen in die Umsetzung des EXMARaLDA-Systems eingeflossen sind bzw. noch einfließen sollen. Dabei wird bewusst auf eine detaillierte Darstellung der einzelnen Systemkomponenten verzichtet, denn die Betonung in diesem Aufsatz soll ja nicht auf den Besonderheiten dieses spe-

ziellen Systems, sondern auf allgemeineren, grundlegenden Aspekten der Gesprächstranskription auf dem Computer liegen.

In diesem Sinne versteht sich auch das Beispiel in Abschnitt 6 weniger als Illustration des EXMARaLDA-Systems im Besonderen denn als Demonstration der Möglichkeiten, die sich durch ein geeignetes solches System im Allgemeinen ergeben können. Der Ausgangspunkt ist dabei ein (hoffentlich) wohlbekanntes Transkript, das nach den GAT-Konventionen angefertigt wurde. Die aufgezeigten Verwendungsmöglichkeiten für dieses Transkript sind teilweise mit den bereits vorhanden Komponenten des EXMARaLDA-Systems zu realisieren. Teilweise deuten sie aber lediglich an, welche Verarbeitungsweisen sich grundsätzlich aufgrund der gewählten System-Architektur für EXMARaLDA-Daten anbieten. Der vorliegende Aufsatz verbindet mit der Andeutung solcher Möglichkeiten nicht zuletzt die Hoffnung, ein größeres Interesse am Thema der Gesprächstranskription auf dem Computer zu wecken und so eine Basis für die Entwicklung des Systems zu finden, die in einem angemessenen Verhältnis zur Größe ihres potentiellen Benutzerkreises steht. Mit anderen Worten: die Qualität eines Systems zur Gesprächstranskription auf dem Computer wächst mit der Anzahl der Personen, die – in welcher Form auch immer – an seiner Entwicklung beteiligt sind, und der Autor möchte die Leserinnen und Leser dieses Aufsatzes ausdrücklich dazu auffordern, von den Möglichkeiten der Kritik und Rückmeldung, die speziell die Veröffentlichung in einem Online-Magazin bietet, Gebrauch zu machen.

1. Warum Gesprächstranskription auf dem Computer?

Die Bedeutung der Transkription natürlicher Gespräche für viele linguistische Forschungszwecke kann wohl als unbestritten angesehen werden: Erst durch das Fixieren des flüchtigen Eindrucks, den die Aufzeichnung eines Gespräches (die ihrerseits bereits eine erste Fixierung des noch flüchtigeren Gesprächsereignisses darstellt) hinterlässt, werden detaillierte linguistische Analysen möglich. Ebenso unbestritten ist wohl die Tatsache, dass durch den Prozess der Transkription der Untersuchungsgegenstand einerseits eine Reduktion erfährt (weil sich nicht alle möglicherweise relevanten Aspekte eines Gespräches verschriftlichen lassen), andererseits jede Transkription bereits eine Interpretation des Transkribierten und damit auch eine erste rudimentäre Analyse darstellt.

Auf die theoretische und praktische Bedeutung der Transkription als solcher für eine wie auch immer geartete Analyse natürlicher Gespräche soll deshalb hier nicht näher eingegangen werden. Vielmehr soll in diesem Abschnitt dargelegt werden, welche Bedeutung der Umstand, dass diese Transkription seit einiger Zeit in der Regel *mit Hilfe des Computers* vorgenommen wird, für die sprachwissenschaftliche Theorie und Praxis hat.

Ursprünglich waren es wohl hauptsächlich praktische Überlegungen, die dazu führten, dass bei der Gesprächstranskription der Rechner nach und nach Bleistift (bzw. Schreibmaschine) und Papier ablöste: einerseits ist der Transkriptionsprozess durch eine ständige Korrektur und Verfeinerung des Transkribierten gekennzeichnet, und solche Korrekturvorgänge sind am Bildschirm wesentlich einfacher und ökonomischer zu bewerkstelligen als auf dem Papier. Ehlich (1992:52f.) drückt dies wie folgt aus:

Revisionsfähigkeit ist also eine wichtige Voraussetzung für ein sachangemessenes Transkriptionsverfahren. [...] Unaufwendige Korrekturmöglichkeiten sind vielleicht das Plus, wenn man den Computer als Textverarbeitungsgerät einsetzt.

Andererseits ergeben sich durch den Umstand, dass das Transkript in digitaler Form vorliegt, wesentlich einfachere und kostengünstigere Möglichkeiten zu seiner Vervielfältigung und Verteilung.

Neben diesen rein praktischen Vorzügen hat der Einsatz des Computers zur Gesprächstranskription aber auch weitergehende methodologische oder sogar theoretische Auswirkungen auf die Art und Weise, wie Gespräche bzw. deren Transkriptionen analysiert werden können. Diese Auswirkungen lassen sich unter drei Schlagworte zusammenfassen:

a) *(teil-)automatisierte Analyse* – Durch Suchfunktionen lassen sich größere Datenmengen gezielt nach bestimmten Phänomenen durchsuchen. Das Ergebnis einer solchen Suche kann sowohl Grundlage einer qualitativen Analyse des jeweiligen Phänomens (d.h. die gefundenen Vorkommen dienen als Beispiele) als auch Basis einer quantitativen Analyse (d.h. die Vorkommen werden gezählt und zu anderen Größen in Bezug gesetzt) sein. Gewisse umfangreiche Formen der empirischen Analyse von Sprachdaten werden erst durch eine solche Automatisierung ermöglicht. Weiterhin lässt sich aber nicht nur die Auswertung, sondern auch bereits die Erstellung gewisser Teile von Transkriptionen teilweise automatisieren. Mechanismen, die bereits erfolgreich auf geschriebene Sprache angewendet werden (z.B. automatische Lemmatisierung, automatisches Taggen), sind im Prinzip auch für Transkriptionen gesprochener Sprache geeignet, und es ist zu erwarten, dass sie auch auf diesem Gebiet zunehmend zum Einsatz kommen werden.

b) *Multi-Media* – Wie bereits oben kurz angesprochen, ist es unbestritten, dass bei der Transkription eines Gespräches zahlreiche Informationen, die in der Audio- oder Video-Aufnahme vorhanden sind, verloren gehen, einfach weil bei der Verschriftlichung der kontinuierliche, analoge Fluss der Ereignisse auf diskrete, digitale Einheiten (d.h. Schriftzeichen und Zeichenketten) abgebildet werden muss. Bei dem sich hieraus ergebenden Dilemma – einerseits wird eine schriftliche Transkription für eine detaillierte Analyse eines Gespräches benötigt, andererseits gehen gerade durch den Prozess des Transkribierens wichtige Details des Gespräches verloren – erweist sich der Einsatz des Computers ebenfalls als hilfreich, indem er die scharfe Abgrenzung zwischen Text einerseits und Bild bzw. Ton andererseits aufhebt. Die hierfür notwendigen Multi-Media-Technologien sind mittlerweile Bestandteil jedes durchschnittlich ausgerüsteten Rechners. Der transkribierte Text kann so nicht nur mit den Ton- oder Videoaufnahmen, die ihm zugrunde liegen, verknüpft werden, sondern es können sogar zusätzliche, nicht-schriftliche Elemente hinzugefügt werden, die eine weitere Beschreibung des Gespräches vornehmen (man denke z.B. an ein Diagramm, das den Tonhöhenverlauf einer Äußerung beschreibt oder eine Grafik, die eine Sitzordnung beschreibt). Eine "Warnung" wie die von MacWhinney (1995:12)

[...] whenever computerized analyses are based not on the original audio signal but on transcribed orthographic forms, one must continue to understand the limits of transcription conventions.

verliert damit unter Umständen an Bedeutung, denn die Vorzüge einer Verschriftlichung und die einer Originalaufnahme können bei einer computergestützten

Verarbeitung von Gesprächstranskriptionen kombiniert und ihre jeweiligen Nachteile damit ausgeglichen werden.

c) *Hypertext* – Die vielleicht bislang am wenigsten beachtete Möglichkeit, die sich erst durch den Einsatz des Computers zur Gesprächstranskription ergibt, ist die Erschließung neuer, nicht-linearer Textformen für die Darstellung von Transkriptionen. Wie Knowles (1995:208) sagt, sind die meisten gängigen Transkriptionsverfahren immer noch an herkömmlichen, auf dem Buchdruck basierenden Textformen orientiert:

The new opportunities are not yet being fully recognized and exploited by linguists [...] Texts are still seen as objects in book format, with words running in horizontal lines from left to right. Annotations are added to these horizontal lines. But book format is an attribute not of speech, but of Western writing systems. There is no reason beyond established custom and practice to present speech in this way. On the contrary, since there are often several annotations relating to the same piece of data, book format is in many cases inappropriate. The use of book format without consideration of other possibilities is based on a confusion between the organization of the data itself, and the presentation of the data on the printed page.

Diese Textformen sind aber ihrer Entstehung nach zweifellos für die *geschriebene* Sprache konzipiert (oder sogar umgekehrt: die geschriebene Sprache ist bezüglich ihrer Organisation auf die Darstellungsart auf einer gedruckten Buchseite optimiert). Natürliche Gespräche weisen hingegen viele Eigenschaften auf, die sich nicht auf Anheb in diese Form abbilden lassen. Insbesondere existieren in der *gesprochenen* Sprache neben sequentiellen Strukturen auch parallele Strukturen (Simultaneität), deren Darstellung in einem prinzipiell linear organisierten Text naturgemäß Probleme bereitet. Gleiches gilt für Annotationen, die sich auf das Transkribierte beziehen, und die Transkription so ebenfalls um eine zusätzliche Dimension erweitern. Zwar stellen nahezu alle Transkriptionssysteme Lösungen für diese Probleme zur Verfügung – meist indem sie Transkripte statt in einzelnen *Zeilen* in *Zeilenbündeln* organisieren.¹ Vom Prinzip der linearen Organisation, das für die Darstellung auf einer gedruckten Seite unabdingbar ist, weichen sie aber in der Regel nicht ab. Für eine Darstellung auf einem Computer-Bildschirm müssen jedoch solche Beschränkungen eigentlich nicht gelten. Die Hypertext-Technologie, die inzwischen so weit verbreitet ist, dass sie bereits als eigene Textkategorie untersucht wird (vgl. z.B. Lobin 1999b, Storrer 1999 oder Bucher 1999), ist in der Lage, durch Verknüpfen einzelner, autonomer Textbestandteile komplexere als lineare Strukturen zu beschreiben und für den Nutzer navigierbar zu machen. Trotz ihres offensichtlichen zusätzlichen Potentials wird diese Möglichkeit für die Darstellung von Transkriptionen gesprochener Sprache zwar bislang wenig genutzt, es ist aber zu erwarten, dass dies in naher Zukunft mehr und mehr der Fall sein wird. Auch in diesem Zusammenhang ergeben sich also durch den Einsatz des Computers zusätzliche Möglichkeiten für die Gesprächstranskription.

¹ Im Gegensatz zu herkömmlichen "eindimensionalen" Texten ist der so entstehende Text damit immerhin zweidimensional. Dies gilt keineswegs nur für Transkriptionssysteme, die die Partiturschreibweise benutzen, sondern z.B. auch für Systeme wie CHAT (MacWhinney 1995), die primär eher an einer vertikalen, also eindimensionalen Darstellung (vgl. Edwards 1993:10f.) orientiert sind.

2. Schwächen existierender Systeme zur Gesprächstranskription auf dem Computer²

Betrachtet man die gängige Praxis der Gesprächstranskription auf dem Computer, stellt man schnell fest, dass man es mit einer fast unüberschaubaren Vielfalt von Systemen, Werkzeugen und Formaten zu tun hat, die auf den ersten Blick wenig gemein zu haben scheinen. Mit den Worten von Bird/Lieberman (2001:24):

While the utility of existing tools, formats and databases is unquestionable, their sheer variety – and the lack of standards able to mediate among them – is becoming a critical problem. Particular bodies of data are created with particular needs in mind, using formats and tools tailored to those needs, based on the resources and practices of the community involved. Once created, a linguistic database may subsequently be used for a variety of unforeseen purposes, both inside and outside the community that created it. Adapting existing software for creation, update, indexing, search and display of 'foreign' databases typically requires extensive re-engineering. Working across a set of databases requires repeated adaptations of this kind.

Kaum eines der vorhandenen Systeme nimmt für sich in Anspruch, in dem Sinne universell zu sein, dass es auch außerhalb des ursprünglich anvisierten Verwendungszweckes zur Anwendung kommen kann. Vielmehr ist fast ausnahmslos zu beobachten, dass die Systeme bereits von ihrer Konzeption her gewisse theoretische Annahmen und Zielsetzungen reflektieren und damit auf eine eng begrenzte Verwendungsweise beschränkt bleiben. Darüber hinaus lässt sich sogar feststellen, dass manche Systeme auch über die technischen Voraussetzungen, die zu ihrem Einsatz nötig sind, ihren potentiellen Benutzerkreis stark eingrenzen.

Wie Bird/Lieberman richtig bemerken, sind die zahlreichen möglichen Verwendungsweisen einer Gesprächstranskription aber im Augenblick ihrer Erstellung keineswegs vorhersehbar. Der beträchtliche finanzielle und zeitliche Aufwand, den eine solche Erstellung erfordert, rechtfertigt es aber, auch diese nur potentiell vorhandenen Möglichkeiten zu berücksichtigen. Zwei Beispiele sollen dies verdeutlichen.

² Um etwaige Missverständnisse auszuschließen sei hier ausdrücklich angemerkt, dass der Begriff "Schwächen" in diesem Zusammenhang in erster Linie diejenigen Aspekte bezeichnet, die bei den angesprochenen Systemen nicht berücksichtigt wurden und aufgrund derer die Systeme in ihrer Verwendungsweise eingeschränkt sind. Dabei muss selbstverständlich beachtet werden, dass zum einen diese Aspekte teilweise bewusst und ausdrücklich aus der Entwicklung der Systeme ausgeklammert wurden, und dass andererseits gewisse Aspekte zum Zeitpunkt ihrer Entwicklung weder theoretisch vorhersehbar noch praktisch umsetzbar waren.

2.1. syncWriter

syncWriter (Rehbein et al. 1993) ist ein Editor für interlinearen Text und wird als solcher zum Anfertigen von Gesprächstranskriptionen in Partiturschreibweise, insbesondere nach den Konventionen von HIAT (Ehlich/Rehbein 1976), benutzt. Das folgende Beispiel stammt aus Rehbein et al. (1993:8):³

| | |
|--|---------------|
| Aeg: Wieviel haben Sie noch von den Fünfigern? | ((sucht etwas |
| Paa: .Ja, nich mehr allzu | |

| |
|---|
| Aeg: in der Schublade, blättert in der Aktenkartei)) |
| Paa: viel. Das heißt, wann unser nächster Termin ist, . das ist die |

| | |
|---|----------|
| Aeg: | [Ah, das |
| Paa: entscheidende Frage. | |
| Sch: ((tritt ein)) Darf ich Ihnen erst mal die Kopie geben? | |

[leise

syncWriter erlaubt die komfortable Eingabe eines solchen Transkriptes in einem Endlosband und stellt Funktionen zur Verfügung, das Transkript (wie hier) auf eine vorgegebene Seitenbreite umzubrechen, um einen Ausdruck zu ermöglichen. Ein Export in Äußerungslisten ist – bei Verlust der Synchronizitätsinformation – ebenfalls vom Programm aus möglich. Weitere vorstellbare Verwendungsweisen einer mit syncWriter erstellten Transkription wären jedoch zum Beispiel:

- eine *vollständige* Überführung der Transkription in eine vertikale Darstellungsform (d.h. ohne Verlust der Synchronizitätsinformation), etwa:

```
Paa: die entscheidende [Frage.]
Sch: [Darf ] ich Ihnen erst mal die Kopie geben?
Aeg: (leise) Ah, das
```

- die Weiterbearbeitung der Transkription in einem Textverarbeitungsprogramm (z.B. Word) oder mit einem anderen Editor (z.B. HIAT-DOS (Ehlich 1992)). Beides wäre vor allem deshalb interessant, weil die Verwendung des syncWriters auf Rechner der Macintosh-Linie beschränkt bleibt - für andere Betriebssysteme existiert keine Version des Programms.
- der Export der Transkription in eine (z.B. relationale) Datenbank, die eine größere Anzahl von Transkriptionen zusammenfasst und für verschiedene Formen der automatischen Analyse – z.B. automatische Suche oder statistische Analyse – zugänglich macht.⁴

Dies scheidet jedoch jeweils daran, dass einerseits das Programm entsprechende Funktionen, also für eine Konvertierung in ein vertikales Format oder geeignete

³ Das Beispiel wird hier nur in seinen Grundzügen und nicht in seiner genauen graphischen Erscheinungsweise reproduziert. Die Verwendung eines äquidistanten Fonts geschieht z.B. hier nur aus layout-technischen Gründen und ist nicht etwa eine vom syncWriter-System vorgegebene Beschränkung.

⁴ Die Datenbank HADAGES (Hamburger Datenbank Gesprochener Sprache) an der Universität Hamburg enthält zwar unter anderem auch aus syncWriter-Transkriptionen exportierte Daten, der Export erfolgt dabei aber über die bereits erwähnten Äußerungslisten, in denen nicht die gesamte in der Transkription vorhandene Information repräsentiert ist.

Export-Funktionen, nicht beinhaltet. Andererseits liegen die Daten in binärer Form vor, so dass sie für andere Anwendungen, die diesen Mangel unter Umständen beheben könnten, nicht zugänglich sind. syncWriter-Daten sind also prinzipiell nur mit dem syncWriter-Programm bearbeitbar und damit *software- und betriebssystemabhängig*.

2.2. GAT

Das Gesprächsanalytische Transkriptionssystem GAT (Selting et al. 1998) orientiert sich an einer vertikalen Darstellung (vgl. Edwards 1993:10f.) von Gesprächstranskriptionen und ist – wie der Name sagt – vornehmlich auf gesprächsanalytische Zwecke ausgerichtet. Es ist eigentlich ausdrücklich (noch) nicht als Computersystem konzipiert:

Die Konventionen richten sich an Transkribenten, nicht an Computer; sie umfassen bisher noch keine Kriterien für die Formatierung von Dateien, die GAT folgend angefertigt werden (Selting et al. 1998:92).

Trotzdem wird an einigen Vorgaben (wie z.B. der Empfehlung, bei Zuordnungsproblemen auf das Darstellungsmittel verschiedener Zeilenabstände zurückzugreifen (ebenda:98)) und Anmerkungen (z.B. "GAT ist jedoch mit jedem Textverarbeitungsprogramm verwendbar" (ebenda:92)) deutlich, dass GAT-Transkriptionen in der Regel auf dem Computer angefertigt werden und daher in digitaler Form vorliegen. Wenn also zwar die Möglichkeit einer automatischen Verarbeitung im Konzept des Systems noch nicht berücksichtigt ist, so ist trotzdem vorstellbar, dass auch GAT-Transkriptionen in einer anderen als der ursprünglich vorgesehen Form auf dem Computer weiterverwendet werden könnten.

So ist es denkbar, dass beim folgenden Beispiel-Ausschnitt (ebenda:116f.)

```

37      S2: ja; V:OLLkommen Widerlich. =ne, (--)
38      un:dh (--) dann hatte er do son (.) son KLEInen BA:RT
39      hier,=ne, (.)
40      und ham wir immer gesagt HIT[ler;=ne,
41      S1:                                     [HITler;
42      S1: h[m,
43      S2: [hm,
```

es für gewisse Zwecke hilfreich sein kann, das Transkript in einer anderen Form darzustellen, z.B. als Partitur,

| |
|---|
| <pre> S2: BA:RT hier,=ne, (.)und ham wir immer gesagt HITler;=ne, hm, S1: HITler; hm,</pre> |
|---|

oder dass so erstellte Transkriptionen in eine Diskurs-Datenbank integriert werden, wie sie am Institut für Deutsche Sprache existiert (Bodmer et al. 2000).

Da das System als Ganzes jedoch noch nicht auf eine computergestützte Verarbeitung der Daten ausgerichtet ist, ist es auch hier schwer bis unmöglich, vorhandene GAT-Daten in dieser Art und Weise weiterzuverwenden: dadurch, dass keine genauen Vorgaben bezüglich der computerseitigen Kodierung von GAT-Transkriptionen existieren, existiert auch keine geeignete Grundlage für die Er-

stellung entsprechender Tools, die eine solche Umrechnung in eine andere Darstellung oder den Export in eine Datenbank bewerkstelligen könnten.

3. Prinzipien für die Gesprächstranskription auf dem Computer

Aus den oben angeführten Beispielen wird deutlich, dass viele existierende Computer-Systeme Transkriptions-Daten produzieren, die sich gewissen potentiellen Verwendungsweisen möglicherweise unnötig verschließen, weil die Systeme diese Verwendungsweisen in ihrer Konzeption nicht berücksichtigen. In diesem Abschnitt soll gezeigt werden, dass und wie sich diese Schwäche durch die Beachtung zweier grundlegender Prinzipien vermeiden lässt.

3.1. Trennung von Inhalt und Darstellung

Es wurde an den oben angeführten Beispielen schon deutlich, dass es für ein und dasselbe Transkriptionsdatum im Prinzip unterschiedliche Darstellungsformen geben kann. Edwards (1993) klassifiziert verschiedene Darstellungstypen für Gesprächstranskriptionen und zeigt, wann und warum eine Form für gewisse Zwecke besser geeignet sein kann als eine andere.

Der Bedarf, ein und dasselbe Dokument auf unterschiedliche Weise darstellen zu können, findet sich aber keineswegs nur im Bereich der Gesprächstranskription. Wie Witt (1999:126ff.) und Lobin (1999b:158ff.) ausführen, kann es sogar als allgemein akzeptierte Erkenntnis der Texttechnologie im weiteren Sinne angesehen werden, dass es möglich und sinnvoll ist, die inhaltliche (logische) Struktur eines Dokumentes von seiner (optischen) Darstellung zu trennen. Verschiedene Darstellungen können dann vom Computer aus einer geeigneten Kodierung des Inhaltes *berechnet* werden. Alle modernen Auszeichnungssprachen wie SGML, XML und HTML verdanken ihre Entstehung dieser Idee, und ihr weit verbreiteter Einsatz in der Computer-Welt zeigt, dass diese auch in die Praxis umzusetzen ist.

Wie Johannsson (1995:82) sagt,

[...] there is far more variation in display than in the features which the display is intended to represent.

kann diese Trennung von Darstellung ("display") und Inhalt ("features which the display is intended to represent") für das Gebiet der Diskurstranskription nicht nur übernommen werden, sondern hat gleichzeitig den weiteren Vorzug, dass Systeme, die auf dem einen oder anderen Darstellungstyp basieren, auf eine gemeinsame Basis gestellt werden können. Um eine Kompatibilität solcher Systeme zu erreichen, genügt es dann, die Inhaltskodierung der Transkription in die jeweils bevorzugte Darstellungsform umzurechnen.

Die Kodierung der inhaltlichen Struktur einer Diskurstranskription ist aber nicht nur insofern wichtig, als sie eine flexible, an die Bedürfnisse verschiedener Benutzer angepasste Darstellung der Daten ermöglicht, sondern sie ist auch für alle anderen Arten der computergestützten Verarbeitung von Transkriptionsdaten unerlässlich.

Bei vielen existierenden Systemen, insbesondere solchen, die hauptsächlich auf eine adäquate Darstellung von Transkriptionen für den menschlichen Benutzer

ausgelegt sind, wird nämlich nicht berücksichtigt, dass nur *implizit* in den verwendeten Zeichenketten vorhandene Informationen für den Computer nicht verwertbar sind. So kann der menschliche Betrachter des Beispiels aus Abschnitt 2.1. aufgrund gewisser Lesegewohnheiten und –erwartungen zwar durchaus erkennen, dass z.B. einerseits die Zeichenkette "Wieviel haben Sie noch von den Fünfzigern?" eine abgeschlossene Äußerung darstellt, die aus mehreren Wörtern besteht, und andererseits die Zeichenkette "((tritt ein))" keine verbale, sondern eine non-verbale Handlung beschreibt. Wenn jedoch ein Computer z.B. eine Wortliste mit allen in diesem Transkriptionsausschnitt gesprochenen Wörtern (und nur diesen) erstellen soll, ist er darauf angewiesen, dass in der Datenstruktur *explizit* vermerkt ist, welche Teile einer Zeichenkette ein Wort darstellen und welche Zeichenketten non-verbale Handlungen beschreiben, also keine geäußerten Wörter sind. Auch für diese und weitere Arten der computergestützten Verarbeitung ist eine Trennung von Inhalt und Darstellung also sinnvoll und notwendig.

3.2. Verwendung von Standards

Viele Aspekte der Inkompatibilität verschiedener Transkriptionssysteme ergeben sich einfach aus dem Umstand, dass diese bereits vor mehreren Jahren erstellt wurden und anschließend nicht an die schnell fortschreitenden Entwicklungen der Computer-Technologie angepasst wurden. Auch dieses Problem ist in keiner Weise auf das Gebiet der Gesprächstranskription beschränkt. Vielmehr lässt sich in nahezu allen Bereichen der elektronischen Datenverarbeitung feststellen, dass die technische Weiterentwicklung nicht nur einen Vorteil darstellt, sondern auch den Nachteil in sich birgt, dass ältere Datenbestände schnell unbrauchbar werden. Vor allem im Kontext der fortschreitenden Vernetzung verschiedener Rechnersysteme haben sich daher in den letzten Jahren in allen Bereichen Initiativen zur Entwicklung von Standards und Empfehlungen herausgebildet, die eine längere Lebensdauer und die Möglichkeit einer breiteren Verwendung von elektronisch kodierten Daten gewährleisten sollen. Soweit möglich, sollte auch ein Computersystem zur Gesprächstranskription von solchen Standards und Empfehlungen Gebrauch machen. Dabei sind einerseits Standards relevant, die die Kodierung elektronischer Dokumente im Allgemeinen betreffen, andererseits Standards, die speziellere Vorgaben zur Kodierung der Dokumentsorte "Gesprächstranskription" machen.

Für erstere stellt sich die Situation relativ überschaubar dar: UNICODE (siehe [UNICODE]) ist mittlerweile der allgemein akzeptierte Standard zur *Kodierung einzelner Zeichen*. Alternative Standards wie der ältere ASCII-Standard kommen für ein generisches System zur Gesprächstranskription nicht ernsthaft in Betracht, da ihr Zeichenvorrat von 256 Zeichen nicht ausreicht, um mehrere Schriftsysteme in *einem* Dokument kodieren zu können, was es beispielsweise bereits unmöglich machen würde, das lateinische und das phonetische Alphabet nebeneinander zu verwenden.

Für die *Kodierung strukturierter Dokumente* bieten sich einerseits der SGML-Standard (siehe [SGML]) und andererseits der auf ihm basierende XML-Standard (der streng genommen erst den Status einer Empfehlung an das W3-Konsortium

besitzt, siehe [XML]) an.⁵ XML wird oft als eine vereinfachte Version von SGML betrachtet und erfreut sich in neuester Zeit einer rapide wachsenden Beliebtheit für alle denkbaren Anwendungen in der Dokument-Technologie. Es zeichnet sich ab, dass XML in den kommenden Jahren *der* Standard zum Datenaustausch über das World Wide Web werden wird. Die angestrebte Langlebigkeit von XML-kodierten Daten scheint damit ebenso gesichert wie eine wachsende Software-Unterstützung auf verschiedensten Systemen. Vor allem letztere war für SGML-kodierte Daten eigentlich nie gegeben – XML ist deshalb wahrscheinlich dem mächtigeren SGML vorzuziehen, wenn es um die Wahl eines Standards zur Kodierung von Gesprächstranskriptionen als strukturierten Dokumenten geht.

Wesentlich unüberschaubarer stellt sich die Situation der Standardisierung von Gesprächstranskription als speziellem Dokumenttyp dar: Im Prinzip kann bereits jede Transkriptionskonvention (wie GAT, CHAT, HIAT etc.) als eine Standardisierungs-Bemühung aufgefasst werden. Es zeichnet sich jedoch in keiner Weise ab, dass eine dieser Konventionen sich gegenüber den anderen durchsetzen würde. Dies liegt wohl einfach darin begründet, dass schon diese Konventionen, wie auch die eventuell zugehörigen Computersysteme, auf bestimmte theoretische Annahmen und Fragestellungen ausgerichtet sind. Eine Standard, der konkret vorschreibt, wie Gesprochenes verschriftlicht wird, ist deshalb in absehbarer Zeit nicht zu erwarten. Standardisierungsbemühungen im Bereich der Gesprächstranskriptionen müssen sich daher darauf beschränken, festzulegen, wie die Elemente einer Transkription, also z.B. Sprecher, Verschriftlichung von Verbalem, Non-Verbalem, etc., organisiert, d.h. in einem computerlesbaren Dokument kodiert werden, ohne dabei näher bestimmen zu können, wie diese Elemente konkret aussehen.

Ein solcher Standard ist das von der Text Encoding Initiative ausgearbeitete *base tag set for transcriptions of spoken language* (siehe z.B. Burnard 1995 und Johansson 1995). Im Rahmen eines Vergleiches verschiedener Transkriptions- und Annotationsschemata bemerken Bird/Lieberman (2001:28) dazu:

The TEI guidelines for 'Transcriptions of Speech' [...] offer access to a very broad range of representational techniques drawn from other aspects of the TEI specification. The TEI report sketches or alludes to a correspondingly wide range of possible issues in speech annotation. All of these seem to be encompassed within our proposed framework [i.e. Annotation Graphs, T.S.], but it does not seem appropriate to speculate at much greater length about this, given that this portion of the TEI guidelines does not seem to have been used in any published transcriptions to date.

Obwohl die TEI-Richtlinien für die Kodierung von Transkriptionen gesprochener Sprache bis vor kurzem der einzige Versuch waren, einen universellen Standard auf diesem Gebiet zu etablieren, scheinen sie sich also nicht durchgesetzt zu haben. Mögliche Gründe hierfür sind einerseits die Festlegung auf bestimmte linguistische Einheiten (das wichtigste Strukturelement in den Richtlinien ist die *utterance*, die keineswegs eine in allen linguistischen Theorien zur Gesprächsforschung vorhandene Kategorie darstellt), andererseits das Fehlen geeigneter Werk-

⁵ Nimmt man von der Idee Abstand, dass Transkriptionen überhaupt in erster Linie Dokumente, d.h. einem Text ähnliche Datenformen sind, sondern betrachtet sie lediglich als Datenstrukturen einer allgemeineren Art, kommt wahrscheinlich auch noch der SQL-Standard für die Kodierung relationaler Datenbanken in Betracht. Vergleiche hierzu z.B. Milde (1999) und Knowles (1995).

zeuge für die benutzergerechte Ein- und Ausgabe von nach diesen Richtlinien kodierten Transkriptionen. Für eine ausführlichere Diskussion dieser Aspekte sei auf Sinclair (1995) verwiesen.

Weitere Standardisierungsbestrebungen existieren, sie befinden sich aber ausnahmslos noch in frühen Stadien der Entwicklung. Wittenburg et al. (2000) beschreiben den EAGLES/ISLE-Standard, der einen Vorschlag für die einheitliche Kodierung von Meta-Informationen über Transkriptionen gesprochener Sprache beinhaltet. Der derzeit vielversprechendste Kandidat für die Standardisierung der Transkriptionen selbst hingegen scheint der von Bird/Lieberman (2001) entwickelte Formalismus der Annotationsgraphen zu sein. Wie weiter unten kurz und in Schmidt (2002) ausführlicher besprochen, liegt dieser Formalismus auch dem hier beschriebenen Transkriptionssystem EXMARaLDA zugrunde.

4. Aspekte der Umsetzung

Die oben angestellten Überlegungen zielen darauf ab, die Voraussetzung für die Entwicklung eines Computersystems zur Gesprächstranskription zu schaffen, das – im besten Falle – einen Großteil der derzeit bestehenden Transkriptionssysteme auf eine gemeinsame Basis stellt und so die Inkompatibilität zwischen diesen Systemen – zumindest bis zu einem gewissen Grad – verschwinden lässt.⁶ Wie bereits oben angesprochen scheint eine wichtige Voraussetzung für ein solches universelles System diejenige zu sein, dass sich die inhaltliche (oder logische) Struktur einer Gesprächstranskription sowohl von ihrer konkreten Darstellung für den menschlichen Benutzer als auch von der Art und Weise ihrer physikalischen Speicherung auf dem Computer trennen lässt. Wie Bird/Lieberman (2001:2) bemerken,

However, we contend that file formats and content specifications are secondary. Instead, we focus on the logical structure of linguistic annotations, since it is here that we observe a striking commonality.

gibt es Anlass zu der Hoffnung, dass sich eine solche logische Struktur nicht nur finden lässt, sondern auch, dass sie für den größten Teil aller Arten von Gesprächstranskriptionen gleich ist.

Der erste Schritt bei der Umsetzung der oben angestellten Überlegungen ist also die Entwicklung eines Modells, das die logische Struktur einer Gesprächstranskription angemessen und in einer für den Computer verwertbaren Form beschreibt. Dabei kann die konkrete Implementierung zunächst zurückgestellt werden.

In Schmidt (2002) wird das EXMARaLDA zugrundeliegende formale Modell zur Beschreibung von Diskurstranskriptionen in seinen Einzelheiten beschrieben. Hier sollen deshalb nur seine wichtigsten Grundzüge aufgelistet werden:

- Ein primärer Zweck einer Gesprächstranskription ist es, das Gespräch in seinem *zeitlichen* Ablauf zu beschreiben. Eine Unterteilung des Gespräches in Ereignisse und eine Zuordnung dieser Ereignisse zu einer geordneten Zeit-

⁶ Faktisch ist es zumindest so, dass EXMARaLDA, da es die Grundlage einer "Datenbank Mehrsprachigkeit" am SFB 538 darstellen wird, zumindest die in dieser Institution gebräuchlichen Transkriptionssysteme, die bereits sehr heterogen sind, vereinigt.

achse muss daher in jedem Falle Bestandteil des Modells sein. Im Gegensatz zur *sprachlichen* Klassifizierung von Gesprächseinheiten hat die zeitliche Einordnung den Vorzug, relativ objektiv, d.h. auch ohne dass irgendwelche sprachtheoretischen Annahmen zugrundegelegt werden, vorgenommen werden zu können. Nicht zuletzt aus diesem Grunde basiert auch das von Bird/Lieberman (2001) beschriebene Annotationsgraphenmodell in erster Linie auf der zeitlichen Struktur von Gesprächstranskriptionen.

- Die eigentliche Verschriftlichung eines Gespräches besteht darin, dass jedes im oben beschriebenen Sinne bestimmte Gesprächs-Ereignis mit Hilfe einer Zeichenkette beschrieben wird. Wie in Abschnitt 1 dargelegt, können auf dem Computer zusätzlich auch andere Beschreibungsarten, wie Bilder oder Tonaufnahmen, benutzt werden. Alle diese Beschreibungsmöglichkeiten werden im System berücksichtigt.
- Fast alle Transkriptionssysteme unterscheiden zusätzlich verschiedene Kategorien von Ereignissen, z.B. verbale Handlungen, non-verbale Handlungen etc. Die verwendeten Kategorien verschiedener Systeme unterscheiden sich dabei so stark, dass es nicht sinnvoll erscheint, vom System her vorzuschreiben, welche Kategorien bei der Transkription verwendet werden sollen. Dennoch muss das System eine Möglichkeit beinhalten, verschiedenartige Ereignisse bezüglich solcher Kategorien klassifizieren zu können.
- Neben der zeitlichen Struktur eines Gespräches beschreibt eine Gesprächstranskription auch immer in irgendeiner Form dessen *sprachliche* Struktur, d.h. es identifiziert sprachliche Einheiten wie Wörter, Äußerungen oder Turns und organisiert diese gegebenenfalls in hierarchischen Strukturen. Auch hier gilt jedoch, dass sich nicht von vorneherein festlegen lässt, welche sprachlichen Strukturelemente verwendet werden, denn diese sind je nach theoretischem Hintergrund und Zielsetzung der Transkription verschieden und in der Regel nicht miteinander vereinbar. Das Modell muss sich also darauf beschränken, eine Möglichkeit zur Verfügung zu stellen, neben der zeitlichen Struktur auch eine nicht näher definierte sprachliche Struktur eines Gespräches beschreiben zu können, ohne sich dabei auf bestimmte Kategorien festzulegen.

Das Annotationsgraphen-Modell stellt für alle diese Aspekte elegante und einfache Lösungen zur Verfügung. EXMARaLDA benutzt daher die Grundzüge dieses Modells, ohne dessen universellen Anspruch zu übernehmen, ein System zur Kodierung jeglicher denkbaren Form der Annotation von Sprachdaten zu sein. Durch die Beschränkung auf den Anwendungsbereich der Transkription natürlicher Gespräche können einerseits einige für die Effektivität der Entwicklung sehr wichtige Vereinfachungen des Modells vorgenommen werden, andererseits bleibt aber gewährleistet, dass EXMARaLDA-Daten zu dem sich unter Umständen herausbildenden Standard des Annotationsgraphen-Modells konform sein werden.

5. EXMARaLDA

Auf der Grundlage der oben angestellten Überlegungen wurde das System EXMARaLDA in einer ersten Version auf dem Computer implementiert. Die zentrale Komponente des Systems sind mehrere Dokumenttypdefinitionen (DTDs), die eine computergerechte XML-Kodierung von Gesprächstranskriptionen festlegen. Diese DTDs setzen die in den Abschnitten 3 und 4 dargelegten Überlegungen um, d.h. sie legen die Syntax eines Dokumentes fest, das eine Gesprächstranskription *inhaltlich* beschreibt. Diese inhaltliche Beschreibung besteht im Wesentlichen aus einer Menge von zeitlichen und sprachlichen Elementen und deren Beziehungen untereinander. Sie ist bezüglich sprachwissenschaftlicher Theorien weitestgehend neutral, so dass verschiedenartige Transkriptionen in einer gemeinsamen Form kodiert werden können.

Die Dokumenttypdefinitionen – und nur diese – bilden die theoretische Basis des Systems. Sie sollen gewährleisten, dass EXMARaLDA-Daten im oben beschriebenen Sinne flexibel und langfristig auf dem Computer bearbeitet werden können. Damit EXMARaLDA-Transkriptionen in der Praxis verwendet oder auch nur erstellt werden können, sind aber weitere Systemkomponenten notwendig. Insbesondere muss dafür gesorgt werden, dass der in einer EXMARaLDA-Datei kodierte Inhalt der Transkription in eine Form gebracht werden kann, die nicht für die Verarbeitung durch den Computer, sondern für die durch einen menschlichen Benutzer optimiert ist. Mit den Worten von Sinclair (1995:107):

Tamás Varádi [vom TELRI-Projekt, T.S.] said quite clearly that a code which is convenient for a machine, which may be essential for a machine, may be extremely awkward for humans. For these new conventions [hier: TEI, T.S.] to gain acceptance in the research community, there will have to be software interpreters provided, that will allow us to move between human and machine without any expenditure of effort whatever. I can't see any alternative to this, and I doubt if anyone who is managing a large spoken resource would want to dispute it. Humans who are going to use transcribed text will want human conventions as much as the transcribers. They will want to see in front of them something that is reasonably readable.

Die hier angesprochene Vermittlung zwischen maschinen- und menschenlesbarem Code betrifft vor allem die Ein- und Ausgabe von Transkriptionen. In seiner derzeitigen Version beinhaltet EXMARaLDA deshalb bereits Software-Tools, die aus einer Inhaltrepräsentation der Daten verschiedene geeignete Darstellungen berechnen können. Weiterhin enthält das System einen Import-Filter, mittels dem Transkriptionen, die unter Einhaltung gewisser minimaler Konventionen⁷ in einem herkömmlichen Text-Editor angefertigt wurden, in eine EXMARaLDA-Transkription überführt werden können. Schließlich wurde ein Prototyp eines Editors erstellt, der das Anfertigen von Transkriptionen in einer Partitur-Oberflä-

⁷ Diese Konventionen betreffen ausnahmslos die "äußere" Organisation von Transkriptionseinheiten, d.h. sie schreiben zum Beispiel vor, dass die Annotation zu einer gegebenen Transkriptionszeile in geschweiften Klammern zu stehen hat, oder dass am Anfang jeder Transkriptionszeile das Sprecherkürzel, gefolgt von einem Doppelpunkt steht. Bezüglich der Verschriftlichung selbst (der "inneren" Organisation) werden keinerlei Vorschriften gemacht. EXMARaLDA öffnet sich auf diese Weise bereits einer großen Zahl vorhandener Transkriptionssysteme, die an einer zeilenweisen Darstellung orientiert sind.

che ermöglicht. In einer kommenden EXMARaLDA-Version wird dieser zu einem gebrauchsfertigen Tool weiterentwickelt werden.⁸

Langfristig sollen EXMARaLDA-Transkriptionen die Basis einer mehrsprachigen Diskursdatenbank bilden. Entsprechende Tools, die nicht auf die Verarbeitung einzelner Dokumente, sondern eher auf das Bündeln und gebündelte Verarbeiten größerer Mengen gleichartiger Transkriptionen abzielen, werden daher ebenfalls zu einem späteren Zeitpunkt in das System integriert werden.

Alle Tools wurden und werden in JAVA programmiert, wodurch ihre Verwendung auf verschiedenen Betriebssystemen möglich ist.⁹ Wie oben ausgeführt, ist EXMARaLDA als System aber auch gar nicht auf diese Tools angewiesen. Vielmehr sieht die System-Architektur vor, dass für spezielle Aufgaben auch spezielle Tools (in prinzipiell beliebigen Programmiersprachen) entwickelt werden können, die aus ihnen resultierenden Daten aber dann wiederum für andere Verwendungsweisen zugänglich sind.

6. Ein Beispiel

Abschließend soll an einem Beispiel deutlich gemacht werden, welche Möglichkeiten der computergestützten Verarbeitung von Gesprächstranskriptionen sich mit dem EXMARaLDA-System ergeben. Als Ausgangspunkt soll dabei wiederum der schon oben verwendete Ausschnitt aus einer Gesprächstranskription aus Selting et al. (1993:116f.) dienen:

| | | |
|----|-----|--|
| 37 | S2: | ja; V:OLLkommen WIderlich. =ne, (--) |
| 38 | | un:dh: (--) dann hatte er do son (.) son KLEInen BA:RT |
| 39 | | hier,=ne, (.) |
| 40 | | und ham wir immer gesagt HIT[ler;=ne, |
| 41 | S1: | [HITler; |
| 42 | S1: | h[m, |
| 43 | S2: | [hm, |

Wie bereits gesagt, macht das hier verwendete Transkriptionssystem GAT keinerlei Annahmen bezüglich der Kodierung dieser Transkription in einer computerlesbaren Datei. Mit dem benutzten Zeichenvorrat ist es aber beispielsweise möglich, diese Transkription, so wie sie oben erscheint, in einer einfachen ASCII-Textdatei zu speichern. Durch die Veränderung einiger Details kann das Transkript so abgeändert werden, dass es zu den Konventionen konform ist, die der EXMARaLDA-Importfilter für solche Text-Dateien vorsieht:

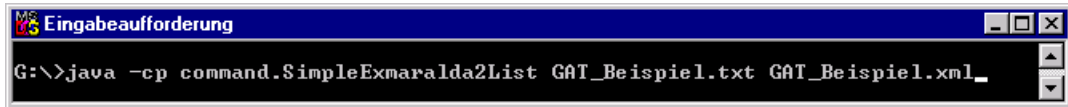
⁸ Nachbemerkung im April 2002: dies ist mittlerweile geschehen, Version 1.1. des Partitur-Editors ist über die Projekt-Homepage verfügbar.

⁹ Leider können aber dennoch nicht alle Betriebssysteme mit der verwendeten JAVA-Version arbeiten – für das Macintosh-Betriebssystem MAC OS 9.x und frühere Versionen existieren nämlich nur veraltete JAVA-Versionen. Macintosh-Nutzer, die EXMARaLDA verwenden wollen, müssten daher auf das neuere MAC OS X umsteigen.

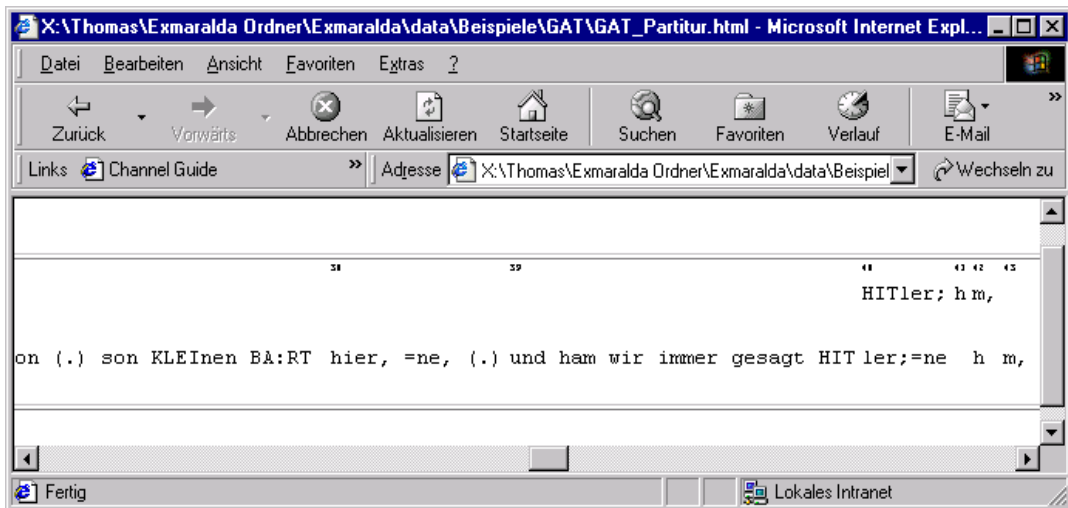
```

S2: ja; V:OLLkommen WIderlich. =ne, (--
S2: un:dh: (--) dann hatte er do son (.) son KLEInen BA:RT
S2: hier,=ne, (.)
S2: und ham wir immer gesagt HIT<ler;=ne,>1>
S1:
    <HITler;>1>
S1: h<m,>2>
S2: <hm,>2>
    
```

Damit ist es möglich, die Datei in ein XML-kodiertes Format zu überführen. In der derzeitigen EXMARaLDA-Version geschieht dies über ein Kommandozeilentool:

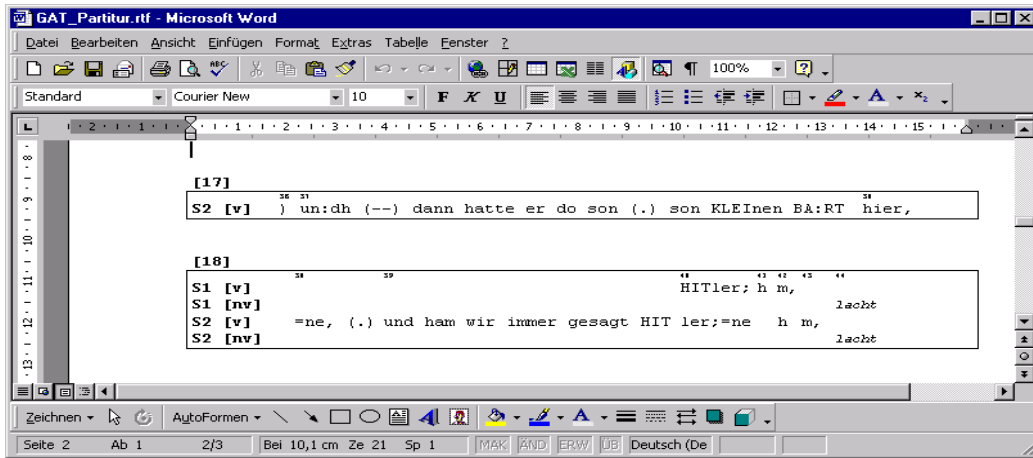


In kommenden Versionen soll ein solcher Import aber komfortabler über einen Dialog in einer graphischen Benutzeroberfläche bewerkstelligt werden können. Da die Transkription jetzt in einem EXMARaLDA-Format vorliegt, können alle auf solchen Dateien operierenden Tools genutzt werden. Beispielsweise kann – wiederum über ein Kommandozeilentool – eine Partiturdarstellung der Transkription berechnet werden und diese dann in einer HTML-Datei kodiert werden. Letztere lässt sich in jedem Internet-Browser (z.B. Internet Explorer oder Netscape) anzeigen. Dadurch ist nicht nur eine Basis für einen Datenaustausch vorhanden, sondern die Transkription kann so auch einfach in Internet-Präsentationen etc. integriert werden:



Alternativ dazu lässt sich die Partitur-Darstellung auch als RTF-Datei kodieren. Dazu ist zunächst ein Umbruch auf eine bestimmte Seitenbreite notwendig. RTF-Dateien können von allen gängigen Textverarbeitungsprogrammen, insbesondere von MS-Word, gelesen werden. Die Transkription kann somit in diesen Programmen mit deren gesamtem Funktionsumfang nach- und weiterbearbeitet (und natürlich auch gedruckt) werden. Beispielsweise ist es so möglich, die Transkrip-

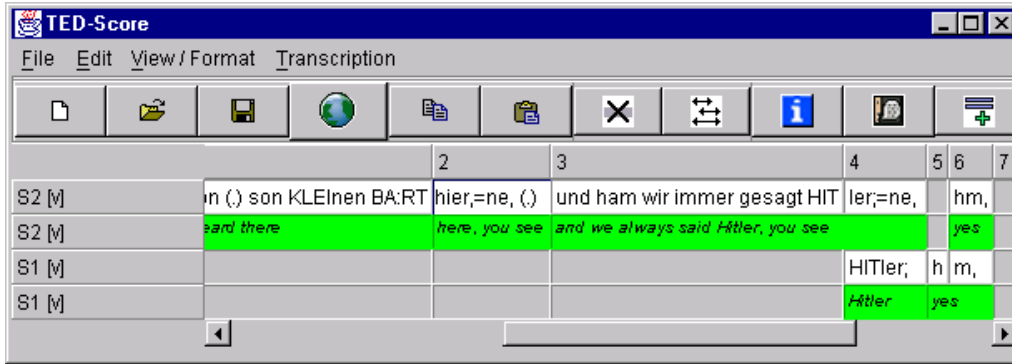
tion ganz oder teilweise per "Copy & Paste" in andere Dokumente, wie z.B. eine Veröffentlichung, einzufügen.



Unter Umständen sollen einer so erstellten Transkription nachträglich weitere Elemente hinzugefügt werden. Beispielsweise kann es für gewisse Verwendungszwecke hilfreich sein, die einzelnen Transkript-Linien mit einer freien englischen Übersetzung zu versehen. In EXMARaLDA kann dies zum Beispiel wiederum in der ursprünglichen Text-Datei geschehen – die Übersetzungen werden dabei einfach den entsprechenden Elementen in geschweiften Klammern nachgestellt:

```
S2: ja; V:OLLkommen Widerlich. =ne, (--
    {yes, absolutely disgusting, isn't it}
S2: un:dh (-- dann hatte er do son (.) son KLEInen BA:RT
    {and then he had a kind of small beard there}
S2: hier,=ne, (.)
    {here, you see}
S2: und ham wir immer gesagt HIT<ler;=ne,>1>
    {and we always said Hitler, you see}
S1:
    <HITler;>1>
    {Hitler}
S1: h<m,>2>
    {yes}
S2: <hm,>2>
    {yes}
```

Möglicherweise komfortabler und übersichtlicher kann derselbe Arbeitsschritt aber auch in einem Partitur-Editor vorgenommen werden:



Die so erstellte Transkription kann jetzt wiederum zusammen mit den nachträglich hinzugefügten Übersetzungen auf verschiedene Arten dargestellt werden. Da die Transkriptionsdaten in einer computerlesbaren Form vorliegen, die Übersetzungen also auch für den Computer als eigene Strukturelemente zu erkennen sind, können sie jedoch ebenso gut bei der Darstellung unberücksichtigt bleiben, ohne dabei aus den Ausgangsdaten gelöscht werden zu müssen.

Die bis hier angeführten Beispiele betreffen lediglich rein praktische Aspekte der Verarbeitung von Gesprächstranskriptionen. Wie im ersten Abschnitt ausgeführt, stellen diese aber nur einen Teil der möglichen Verwendungsweisen dar.

Um die Möglichkeit einer teil-automatisierten Analyse der Transkriptionsdaten zu illustrieren, mag es hilfreich sein einen kleinen Blick auf die computerseitige Kodierung der Beispiel-Transkription zu werfen. Der folgende Ausschnitt stammt aus der XML-Datei die auf die oben beschriebene Art und Weise aus einer Text-Datei importiert und anschließend automatisch segmentiert wurde:

```
[...]
<segmentation name="event">
[...]
<segment id="TIE1.EVT76" start="T36" end="T37">ja; V:OLLkommen WIderlich.=ne, (--
)</segment>
<segment id="TIE1.EVT78" start="T37" end="T38">un:dh (-- ) dann hatte er do son
(.) son KLEInen BA:RT </segment>
<segment id="TIE1.EVT80" start="T38" end="T39">hier, =ne, (.)</segment>
<segment id="TIE1.EVT82" start="T39" end="T40">und ham wir immer gesagt
HIT</segment>
<segment id="TIE1.EVT83" start="T40" end="T41">ler;=ne</segment>
<segment id="TIE1.EVT90" start="T42" end="T43">h</segment>
<segment id="TIE1.EVT91" start="T43" end="T44">m, </segment>
[...]
</segmentation>
<segmentation name="word">
<segment id="TIE1.WRD108" start="T185" end="T186">V:OLLkommen</segment>
<segment id="TIE1.WRD109" start="T186" end="T187">WIderlich.</segment>
[...]
<segment id="TIE1.WRD113" start="T190" end="T191">hatte</segment>
<segment id="TIE1.WRD114" start="T191" end="T192">er</segment>
<segment id="TIE1.WRD115" start="T192" end="T193">do</segment>
<segment id="TIE1.WRD116" start="T193" end="T194">son</segment>
<segment id="TIE1.WRD118" start="T195" end="T196">son</segment>
<segment id="TIE1.WRD119" start="T196" end="T197">KLEInen</segment>
<segment id="TIE1.WRD120" start="T197" end="T38">BA:RT</segment>
[...]
</segmentation>
```

Auch wenn diese Darstellung offensichtlich nicht vornehmlich für den menschlichen Betrachter konzipiert ist, so lässt sich doch erkennen, dass hier ein strukturiertes Dokument vorliegt, das verschiedene kleinere Einheiten (Segmente: "seg-

ment") zu größeren Einheiten (Segmentierungen: "segmentation") zusammenfasst, die jeweils mit verschiedenen Attributen (IDs, Start-/Endpunkten etc.) versehen sind. Die in diesem Ausschnitt enthaltenen Elemente ergeben in ihrer Gesamtheit einerseits die zeitliche Struktur ("event") der Gesprächstranskription, andererseits enthalten sie aber auch sprachliche Strukturelemente ("word"). Eine teil-automatisierte Analyse, beispielsweise eine automatische Suche, wird sich in der Regel auf letztere beziehen. Dabei stellt sich mit den hier benutzten Transkriptionskonventionen unter Umständen das Problem, dass verschiedene Aussprachevarianten ein- und desselben Wortes bei der Verschriftlichung, d.h. bei der Abbildung auf Zeichenketten, verschiedene Formen annehmen. Edwards (1992:446f.) beschreibt unter dem Stichwort "consistent encoding for exhaustive retrieval", warum dies bei einer automatischen Suche zu Problemen führen kann – die für die Transkription verschiedener sprachlicher Realisierungen eines bestimmten Lexems verwendeten Zeichenketten sind möglicherweise nicht vorhersehbar, und es besteht somit die Gefahr, dass bei der Suche bestimmte Formen unbeabsichtigt (aber systematisch) übersehen werden. Da aber gleichzeitig die Verschiedenheit dieser Formen eine durchaus wünschenswerte Detailtreue der Transkription reflektiert, ist es auch nicht akzeptabel, sie einer vereinheitlichten Schreibweise zu "opfern". Liegen die Daten in einer Form wie der oben angedeuteten vor, ist dies jedoch gar nicht nötig, sondern es ist möglich, beide Formen – eine detailgetreue und eine vereinheitlichte – *nebeneinander* zu kodieren. Praktisch kann dies durch Bezugnahme auf die ID-Attribute in den jeweiligen XML-Elementen geschehen, etwa:

| ID | Variante | Lemma |
|-------------|-------------|------------|
| TIE1.WRD108 | V:OLLkommen | vollkommen |
| TIE1.WRD109 | WIderlich | widerlich |
| TIE1.WRD113 | hatte | haben |
| TIE1.WRD114 | er | er |
| TIE1.WRD115 | do | da |
| TIE1.WRD116 | son | so |
| TIE1.WRD116 | son | ein |
| TIE1.WRD118 | son | so |
| TIE1.WRD118 | son | ein |
| TIE1.WRD119 | KLEInen | klein |
| TIE1.WRD120 | BA:RT | Bart |

Die Lemmata sind für eine Darstellung für den menschlichen Benutzer wahrscheinlich in den meisten Fällen uninteressant und können dementsprechend dort auch unberücksichtigt bleiben. Ihr Vorhandensein ermöglicht aber eine automatische Suche, die nicht unter den oben beschriebenen Unsicherheiten leidet. So würde eine Suche nach dem Lemma „und“ im gesamten Transkript unter anderem folgende Transkriptzeilen liefern:

Suche nach Lemma "und":

- S2: und (.) die ältere tochter is AUCh in berlin gewesen?
- S2: und (.) der KE::RL,
- S2: un:dh: bei !JE!der gelegenheit hat er die polizeI
- S2: un wenn da: einmal: jemand zum abschied geHUPT hat,
- S2: =und NICHTS,=
- S2: ja und dann PLÖTZlich (--) is seine FRAU abgehaun;=ne,=
- S2: un:dh (--) dann hatte er do son (.) son KLEInen BA:RT
- S2: und ham wir immer gesagt HITler;=ne,

Obwohl hier nur in einer einzelnen Transkription gesucht wurde, ist die Problematik, die sich aus der Verschiedenheit der Formen ergibt, offensichtlich. Es sollte daher klar sein, dass bei größeren Datenmengen, d.h. z.B. für die Suche auf einem gesamten Korpus, eine computergerechte Aufarbeitung der Daten, wie EXMARaLDA sie ermöglicht, unabdingbar ist.

Das Anreichern der Transkription um zusätzliche Information bleibt natürlich nicht auf solche grundlegenden Aspekte beschränkt. Vielmehr macht es die im EXMARaLDA-Dokument kodierte sprachliche Struktur möglich, beliebige linguistische Annotationen, die sich auf die vorhandenen Strukturelemente beziehen, hinzuzufügen. Beispielsweise ist eine Annotation nach Phrasenstrukturen denkbar,

| ID | Variante | Phrase |
|-------------|-------------|-----------|
| TIE1.WRD118 | ein | NP |
| TIE1.WRD119 | peneTRANter | |
| TIE1.WRD120 | !WI!derling | |
| [...] | [...] | [...] |
| TIE1.WRD118 | son | NP |
| TIE1.WRD119 | KLEInen | |
| TIE1.WRD120 | BA:RT | |

die dann ebenfalls Gegenstand entsprechender Suchanfragen sein kann:

Suche nach Phrase "NP":

- S2: das war aber ein peneTRANter: !WI!derling.=also (1.0)
- [...]
- S2: un:dh (--) dann hatte er do son (.) son KLEInen BA:RT

Die Gesprächstranskription öffnet sich damit unterschiedlichsten Verwendungs- und Analysezielen. Dabei ist durch die Strukturierung der EXMARaLDA-Dokumente immer sichergestellt, dass für den jeweiligen Verwendungszweck unnötige Information unberücksichtigt bleiben kann und somit nicht als störender Ballast bei der Bearbeitung mitgetragen werden muss.

All diese Beispiele illustrieren lediglich *eine* der im ersten Abschnitt angesprochenen Möglichkeiten, die sich durch ein geeignetes System zur Gesprächstranskription auf dem Computer ergeben, nämlich die der (teil-)automatisierten Analyse von Transkriptionsdaten. Die beiden anderen Möglichkeiten –

Multimedia und Hypertext – lassen sich selbst in einem Artikel einer Online-Zeitschrift nur schwer demonstrieren, denn sie arbeiten beide mit Textformen, die von denen des Buchdrucks losgelöst sind. Statt also hier umständlich zu versuchen, die Integration von multimedialen Daten oder eine nicht-lineare Darstellungsform für Transkriptionen auf dem "Papier" anzudeuten, sei auf die EXMARALDA Homepage

<http://www.rrz.uni-hamburg.de/exmaralda/>

verwiesen, auf der solche Beispiele präsentiert werden (das hier verwendete Beispiel findet sich dort unter dem Namen "ein widerlicher Kerl").

7. Literatur

- Bird, Steven / Liberman, Mark (2001): A formal framework for linguistic annotation. In: *Speech Communication* 33(1,2), 23-60.
- Bodmer, Franck / Fach, Marcus / Schmidt, Rudolf / Schütte, Wilfried (2000): Von der Tonbandaufnahme zur integrierten Text-Ton-Datenbank. Instrumente für die Arbeit mit Gesprächskorpora. In: *Tagungsband zur 1. Freiburger Arbeitstagung zur Romanistischen Korpuslinguistik*, Freiburg.
- Bucher, Hans-Jürgen (1999): Die Zeitung als Hypertext. In: Lobin, Henning (Hg.): *Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering*. Wiesbaden: Westdeutscher Verlag, 9-32.
- Burnard, Lou (1995): The Text Encoding Initiative: an overview. In: Leech, Geoffrey / Myers, Greg / Thomas, Jenny (Hg.): *Spoken English on Computer: Transcription, Markup and Application*. Harlow: Longman, 69-81.
- Edwards, Jane / Lampert, Martin (Hg.) (1993): *Talking Data – Transcription and Coding in Discourse Research*. Hillsdale: Erlbaum.
- Edwards Jane (1993): Principles and Contrasting Systems of Discourse Transcription. In: Edwards, Jane / Lampert, Martin (Hg.): *Talking Data – Transcription and Coding in Discourse Research*. Hillsdale: Erlbaum, 3-31.
- Ehlich, Konrad (1993): HIAT - a Transcription System for Discourse Data. In: Edwards, Jane / Lampert, Martin (Hg.): *Talking Data – Transcription and Coding in Discourse Research*. Hillsdale: Erlbaum, 123-148.
- Ehlich, Konrad (1992): Computergestütztes Transkribieren - das Verfahren HIAT-DOS. In: Richter, Günther (Hg.) *Methodische Grundfragen der Erforschung gesprochener Sprache*, Frankfurt a.M.: P. Lang, 47-59
- Ehlich, Konrad / Rehbein, Jochen (1976): Halbinterpretative Arbeitstranskriptionen (HIAT). In: *Linguistische Berichte* 45, 21-41.
- Johansson, Stig (1995): The approach of the Text Encoding Initiative to the encoding of spoken discourse. In: Leech, Geoffrey / Myers, Greg / Thomas, Jenny (Hg.): *Spoken English on Computer: Transcription, Markup and Application*. Harlow: Longman, 82-98.
- Knowles, Gerry (1995): Converting a corpus into a relational database: SEC becomes MARSEC. In: Leech, Geoffrey / Myers, Greg / Thomas, Jenny (Hg.): *Spoken English on Computer: Transcription, Markup and Application*. Harlow: Longman, 208-219.

- Leech, Geoffrey / Myers, Greg / Thomas, Jenny (Hg.) (1995): Spoken English on Computer: Transcription, Markup and Application. Harlow: Longman.
- Lobin, Henning (Hg.) (1999a): Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering. Wiesbaden: Westdeutscher Verlag.
- Lobin, Henning (1999b): Intelligente Dokumente. Linguistische Repräsentation komplexer Inhalte für die hypermediale Wissensvermittlung. In: Lobin, Henning (Hg.): Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering. Wiesbaden: Westdeutscher Verlag, 155-178.
- MacWhinney, Brian (1995): CHAT Manual. (childes.psy.cmu.edu/pdf/chat.pdf)
- Milde, Jan-Torsten (1999): Effizientes Document Engineering sprachlicher Daten. In: Lobin, Henning (Hg.): Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering. Wiesbaden: Westdeutscher Verlag, 197-220.
- Rehbein, Jochen et al. (1993): Manual für das computergestützte Transkribieren mit dem Programm syncWRITER nach dem Verfahren der Halbinterpretativen Arbeitstranskriptionen (HIAT). Universität Hamburg: Germanisches Seminar.
- Schmidt, Thomas (2002): EXMARaLDA – ein System zur Diskurstranskription auf dem Computer. In: Arbeiten zur Mehrsprachigkeit (Working Papers in Multilingualism), Serie B (34). Hamburg
- [SGML] : International Organization for Standardization ISO 8879: Information processing - Text and office systems - Standard Generalized Markup Language (SGML) ISO1986 Annexes.
- Selting, Margret et al. (1998): Gesprächsanalytisches Transkriptionssystem (GAT). In: Linguistische Berichte (173), 91-122.
- Sinclair, John (1995): From Theory to Practice. In: Leech, Geoffrey / Myers, Greg / Thomas, Jenny (Hg.): Spoken English on Computer: Transcription, Markup and Application. Harlow: Longman, 99-109.
- Storrer, Angelika (1999): Kohärenz in Text und Hypertext. In: Lobin, Henning (Hg.): Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering. Wiesbaden: Westdeutscher Verlag, 33-66.
- [UNICODE] : The Unicode Consortium. The Unicode Standard, Version 3.1.1, which consists of: The Unicode Standard, Version 3.0 (Reading, MA, Addison-Wesley, 2000. ISBN 0-201-61633-5), as amended by the Unicode Standard Annex #27: Unicode 3.1 [<http://www.unicode.org/reports/tr27/>] and the Unicode 3.1.1 Update Notice [www.unicode.org/versions/Unicode3.1.1.html].
- Witt, Andreas (1999): SGML und Linguistik. In: Lobin, Henning (Hg.): Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering. Wiesbaden: Westdeutscher Verlag, 155-178.
- Wittenburg, Peter, Broeder, Daan and Sloman, B., (2000), EAGLES/ISLE: A Proposal for a Meta Description Standard for Language Resources, White Paper. LREC 2000 Workshop, Athens.
- [XML] : W3C Recommendation: Extensible Markup Language (XML) 1.0 (2nd edition 6 October 2000) [www.w3c.org]

Thomas Schmidt
Universität Hamburg
SFB 538 "Mehrsprachigkeit"
Max Brauer-Allee 60
22767 Hamburg
thomas.schmidt@uni-hamburg.de

Veröffentlicht am 17.4.2002

© Copyright by GESPRÄCHSFORSCHUNG. Alle Rechte vorbehalten.