

Stellungnahme zu Wolfgang Schneiders Artikel "Annotate in Transkriptionen aus DV-technischer Sicht"

Thomas Schmidt

1. Einleitung

Diese Stellungnahme setzt sich mit Wolfgang Schneiders Artikel "Annotationsstrukturen in Transkripten" auseinander. Als Entwickler des EXMARaLDA-Systems, das Schneider exemplarisch für die in seinem Aufsatz entwickelten Konzepte diskutiert, habe ich mich aufgerufen gefühlt, einige Sachverhalte aus meiner Sicht zu schildern.

Die Versuchung, auf viele der zahlreichen Details, die Schneiders Aufsatz anspricht, eine ebenso detaillierte Antwort zu geben, ist groß; ebenso die Versuchung, aus dieser Stellungnahme eine Abhandlung über die Mühen und Früchte der EXMARaLDA-Entwicklung im Allgemeinen zu machen. Ich habe versucht, beiden Versuchungen zu widerstehen und mich – der Kürze und Klarheit zuliebe – auf einige wenige Aspekte beschränkt, die mir wesentlich erscheinen. Ich hoffe, dass die von Schneider initiierte Diskussion dadurch nicht beendet, sondern angeregt wird.

2. Zusammenfassung und allgemeine Stellungnahme

Im ersten Teil seines Artikels entwickelt der Autor anhand von fünf im deutschsprachigen Raum verbreiteten Transkriptionskonventionen eine Typologie von Annotationsverfahren für gesprächsanalytische Transkriptionen. Das erklärte Ziel ist es, damit ein Bewertungskriterium für computergestützte Transkriptionssysteme zur Verfügung zu stellen – ein solches System ist demnach als umso flexibler zu bewerten, je mehr dieser Verfahren es unterstützt. Im zweiten Teil des Aufsatzes wird dann das EXMARaLDA-System exemplarisch anhand der im ersten Teil entwickelten Kriterien untersucht. Als Fazit wird festgestellt, dass EXMARaLDA zwar einen vielversprechenden und im Sinne der entwickelten Kriterien richtigen Ansatz beinhaltet, diesen aber nicht konsequent genug umsetzt.

Auch wenn sich der Artikel erklärtermaßen auf DV-technische Aspekte der Transkription beschränkt, ist seine Perspektive sicherlich keine einseitige; Schneiders Ausführungen decken in vielerlei Hinsicht ein großes Spektrum ab: sowohl die Auswahl der berücksichtigten Transkriptionskonventionen als auch die in Betracht gezogenen Transkriptionswerkzeuge machen deutlich, dass dem Autor an einer von bestimmten gesprächsanalytischen Schulen unabhängigen Sichtweise des Gegenstandes gelegen ist. Ebenso spricht er Entwickler und Benutzer von computergestützten Transkriptionssystemen gleichermaßen und auf mehreren Ebenen an – die Ausführungen reichen von der Problematik der langfristigen Archivierbarkeit von Transkriptionsdaten über Betrachtungen zu texttechnologischen Mitteln wie Auszeichnungssprachen bis hin zu implementationspezifischen Details des EXMARaLDA-Systems auf der Daten- und Software-Ebene.

Diese Vorgehensweise führt mindestens zu zwei erfreulichen Resultaten: Zum einen demonstriert der Autor, indem er ein und denselben Gegenstand an mehreren, teilweise zueinander in Konkurrenz stehenden Transkriptionssystemen untersucht, dass eine auf formale Aspekte beschränkte Betrachtung dieser Systeme ihre Gemeinsamkeiten und Unterschiede in systematischer Weise offen legen kann. Er leistet damit einen Beitrag zu einer meines Erachtens viel zu selten geführten Diskussion. Zum anderen kann der Aufsatz aufgrund seines breiten Adressatenkreises als Grundlage einer Auseinandersetzung dienen, zu der Schneider in einer Nachbemerkung ausdrücklich selbst aufruft – entscheidende Fortschritte bei der Entwicklung computergestützter Transkriptionssysteme sind nur dann zu erwarten, wenn diese Entwicklung als eine *wissenschaftliche* Aufgabe begriffen wird, an der sich Gesprächsforscher und Informatiker in gleicher Weise und in gegenseitigem Austausch beteiligen. Ich halte diesen Aufruf für wichtig und richtig und möchte ihm mit meiner Stellungnahme folgen.

Meine Kritik bezieht sich – erwartungsgemäß – auf die im zweiten Teil des Aufsatzes entwickelten Ausführungen zum EXMARaLDA-System. Hier zieht Schneider meines Erachtens einige Schlüsse, die im besten Falle nur voreilig oder missverständlich formuliert, im schlimmsten Falle sogar schlicht falsch sind. Auf diese und die Irrtümer, die ihnen zugrunde liegen mögen, werde ich im zweiten Abschnitt im Detail eingehen.

Auf einer allgemeineren Ebene möchte ich hier jedoch zunächst kurz eine Problematik ansprechen, die sich weniger aus dem, was Schneider schreibt, denn aus dem, was er *nicht* schreibt, ergibt: Es ist sicherlich nicht die Absicht des Autors, eine Mängelliste zu EXMARaLDA zu erstellen (genauso wenig soll diese Stellungnahme eine Aufzählung seiner Vorzüge sein). Er legt aber dennoch den Schwerpunkt auf diejenigen Aspekte des Systems, die nach seiner Einschätzung – sei es per Konzept oder "nur" in der Implementierung – dessen Unzulänglichkeiten demonstrieren. Der (ab-)geneigte Leser mag so bei selektiver Lektüre zu dem Schluss kommen, dass EXMARaLDA und der Partitur-Editor wenig können, was andere Transkriptionssysteme mit den zu ihnen gehörenden Computerwerkzeugen nicht können, und dass zudem einige Dinge, die bereits zum *Status quo* gehören, mit EXMARaLDA nicht realisierbar sind. Formulierungen wie

[Ein Teil der Transkribenten wird] auf einen Fortschritt bei der Entwicklung von Transkriptionseditoren hoffen [...], weil sie einen Editor brauchen, der mehr Strukturen für das Annotieren vorsieht und der zu der von ihnen bevorzugten Konvention exakt passt. (228)

könnten den Schluss nahe legen, dass es ökonomischer sei, auf eine Weiter- oder Neuentwicklung von EXMARaLDA zu warten, als sich mit einem unfertigen System auseinander zu setzen, das die in es gesetzten Erwartungen nicht zu erfüllen vermag. Ich würde einen solchen Schluss als sehr bedauerlich empfinden, denn gerade das durchaus ermutigende Interesse an der bisherigen Arbeit¹ und die kritischen Rückmeldungen von Transkribenten haben entscheidend zur Weiterentwicklung von EXMARaLDA beigetragen. Um diese und weitere Transkribenten – übrigens durchaus im Sinne dessen, was Schneider in seiner Schlussbemerkung schreibt – zu einer Fortführung dieser Auseinandersetzung zu ermutigen,

¹ Seit der Veröffentlichung vor etwas mehr als einem Jahr wurden 247 Zugangsberechtigungen für den Download der EXMARaLDA-Software angefordert.

und weil diese Aspekte in Schneiders Artikel meines Erachtens zu kurz kommen, sei an dieser Stelle auf einige Punkte hingewiesen, die demonstrieren, dass EXMARaLDA und der Partitur-Editor bereits in ihrer jetzigen unvollendeten Form einen Fortschritt in vielerlei Hinsicht bedeuten. Den Vergleich mit existierenden Systemen zur computergestützten Transkription überlasse ich dem Leser:

- EXMARaLDA-Transkriptionen sind problemlos zwischen Linux-, Unix-, Windows-, und Macintosh-Systemen austauschbar und können dort jeweils mit den gleichen Werkzeugen bearbeitet werden.
- EXMARaLDA-Transkriptionen können in Zeilenschreibweise (z.B. weitgehend nach GAT) eingegeben und danach automatisch in eine Partiturdarstellung überführt werden.²
- Eine im TASX-Annotator³ am synchronisierten Audio-Signal angefertigte Transkription kann in den Partitur-Editor importiert und dort weiterbearbeitet und als Partitur ausgegeben werden. Anschließend kann sie nach Praat exportiert werden, um dort als Grundlage einer phonetischen Analyse zu dienen. Der umgekehrte Weg (von Praat über EXMARaLDA nach TASX) ist ebenso möglich, wie auch ein Export in das von Bird/Lieberman vorgeschlagene *Atlas Interchange Format* (AIF).
- Eine türkische EXMARaLDA-Transkription kann unter Verwendung der jeweils üblichen Schriftsysteme mit Übersetzungen ins Portugiesische, Russische, Griechische oder Japanische oder mit einer phonetischen Transliteration im IPA versehen werden, und zwar unabhängig von speziell dafür entwickelten Schriftsätzen.
- Unter Einhaltung einiger minimaler Konventionen kann aus einer EXMARaLDA-Transkription eine alphabetische Wortliste generiert werden, die über Hyperlinks mit der eigentlichen Transkription verknüpft ist.
- Die EXMARaLDA-DTDs und der Partitur-Editor sind kostenlos und mit minimalem organisatorischen Aufwand erhältlich.
- Ein Benutzer des EXMARaLDA-Systems kann seine Beobachtungen und Verbesserungsvorschläge dem Entwickler mitteilen, um so Einfluss auf die weitere Entwicklung zu nehmen.
- Für einen Entwickler, der in acht Jahren eine neue Anwendung für EXMARaLDA-Daten programmieren möchte, ist die Struktur dieser Daten ohne größeren Aufwand zu erschließen. Die Wahrscheinlichkeit, dass gängige Programmiersprachen und Entwicklungsumgebungen ihn bei dieser Aufgabe unterstützen (indem sie entsprechende Schnittstellen zum Einlesen und Bearbeiten von XML-Daten zur Verfügung stellen), ist relativ hoch.

² Dass sich mit dieser Methode GAT-Transkripte automatisch in HIAT-Transkripte überführen lassen, ist aber – wie Schneider richtig anmerkt – nicht wahr und wurde von mir auch nie behauptet.

³ Es sei an dieser Stelle angemerkt, dass es angesichts des Funktionsumfangs des TASX-Annotators (von Video-/Audio-Synchronisation über Pausenerkennung bis hin zu verschiedenen Import-/Exportfiltern und Darstellungsformen) etwas befremdlich anmutet, wenn von ihm als einem "einfachen Transkriptionseditor" (215, Fußnote 48) gesprochen wird. Was wäre demnach ein *komplexer* Transkriptionseditor?

- Eine Entwicklerin, die in fünf Jahren ein computergestütztes Transkriptionssystem entwickeln möchte, kann auf öffentlich zugängliche Artikel über EXMARaLDA zurückgreifen, die ihr ersparen, das Rad neu erfinden zu müssen, oder ihr helfen, Fehler, die bei der EXMARaLDA-Entwicklung gemacht wurden, zu vermeiden.

Anders als die Ausführungen im nächsten Abschnitt widerspricht diese (unvollständige) Auflistung nicht unbedingt dem, was Schneider in seinem Artikel schreibt, und sie kann auch nicht dazu dienen, die an EXMARaLDA geübte Kritik in allen Punkten auszuräumen. Dennoch sollte sie deutlich machen, dass der im obigen Zitat erwähnte "Fortschritt" sich mit EXMARaLDA teilweise bereits eingestellt hat, und zwar in einer Form, die es Gesprächsforschern und Informatikern gleichermaßen möglich machen sollte, über weitere Verbesserungen zu diskutieren. Eine solche Diskussion wäre der Sache unbestreitbar dienlicher als ein bloßes "Hoffen" auf weitere Entwicklungen.

3. Populäre Irrtümer

Der von Schneider geforderte Dialog zwischen Gesprächsforschern und Informatikern findet – wenn vielleicht auch weniger oft als wünschenswert – bereits statt. Er gestaltet sich aber oft dort schwierig, wo es weniger um konkrete Anwendungsfragen als um allgemeinere Konzepte geht. Teilweise verantwortlich dafür sind meiner Meinung nach zwei populäre Irrtümer, derer Schneider sich wohl durchaus bewusst ist, von denen er sich in seiner Darstellung aber auch nicht immer ganz frei zu machen weiß. Der erste dieser Irrtümer lautet:

3.1. "EXMARaLDA ist ein Transkriptionseditor"

Varianten dieses Irrtums lauten beispielsweise "EXMARaLDA ist ein Editor für HIAT-Transkriptionen" oder auch "EXMARaLDA wurde am SFB 538 als Nachfolger für das syncWriter-Programm konzipiert" (227).

Wie der Name "*Extensible Markup Language for Discourse Annotation*"⁴ sagt, ist EXMARaLDA aber in erster Linie eben kein *Programm*, sondern eine *Sprache* zur Beschreibung von Diskursannotationen. Der Partitur-Editor versteht sich ausdrücklich als *eines von vielen* denkbaren Ein- und Ausgabewerkzeugen für in dieser Sprache formulierte Transkriptionsdaten. Die grundlegende Idee hinter EXMARaLDA ist die gleiche, die für den Einsatz von Markup-Sprachen im Allgemeinen gilt: wenn statt einer graphischen Darstellung von Daten deren logische Struktur kodiert wird, wird es möglich, mehrere, für bestimmte Zwecke optimierte Darstellungen aus der logischen Struktur zu berechnen, oder, noch allgemeiner, die Daten für verschiedene Zwecke (also nicht unbedingt nur für die graphische Darstellung) flexibel einsetzbar zu machen.

⁴ Zugegebenermaßen ist es dem schöneren Akronym geschuldet, wenn hier von "discourse annotation" gesprochen wird – dies ist nicht weiter theoretisch motiviert, d.h. weder soll hier "Diskurs" in Opposition zu "Konversation" (wie z.B. bei "Diskursanalyse" vs. "Konversationsanalyse"), noch "Annotation" in Opposition zu "Transkription" stehen. Vielmehr ist das Kopfnomen "Language" als zentral zu betrachten. "Extensible Markup Language for Conversation Transcription" wäre also ebenso angemessen.

EXMARaLDA entsteht im Rahmen eines Projektes am SFB 538. Dessen Ziel ist die Erstellung einer Datenbank "Mehrsprachigkeit", in der alle am SFB vorhandenen Transkriptionsdaten in einer Form gebündelt werden, in der sie gemeinsam verarbeitet werden können. Die am SFB verwendeten Transkriptionskonventionen und -programme sind noch weitaus vielfältiger als die von Schneider herangezogenen Systeme;⁵ viele Projekte benutzen weder für die Eingabe noch für die Ausgabe ihrer Daten die Partiturschreibweise, einige Projekte benutzen gar überhaupt keine Ausgabe im Sinne einer geschlossenen Darstellung auf Bildschirm oder Papier. Was die Neuentwicklung eines Partitur-Editors vor allem motiviert hat, ist die Tatsache, dass Daten aus bestehenden Partitur-Editoren (im Gegensatz zum weitaus größten Teil anders gearteter Transkriptionsdaten) sich jeglicher Form der Weiterbearbeitung entziehen.⁶ Der Partitur-Editor ist also keinesfalls die zentrale Architekturkomponente des EXMARaLDA-Systems, sondern lediglich die Eingangs-Schnittstelle für diejenigen Transkribenten, die auf einer Partiturschreibweise bestehen. Desgleichen ist auch das vom Editor benutzte Datenformat nicht das zentrale EXMARaLDA-Format – die Bezeichnung "*basic-transcription*" könnte bereits deutlich machen, dass es sich dabei nur um die Datengrundlage für eine weitere Bearbeitung handelt. Dies ist in Schmidt (2002a) im Detail ausgeführt.

Zu Beginn des zweiten Teiles seines Aufsatzes formuliert Schneider auch noch dementsprechend:

EXMARaLDA ist eine Sammlung von formalen, also nicht an den Inhalt gebundenen, DV-technischen Konventionen zur Kodierung von Transkripten und dazu passenden Werkzeugen. (214)

Leider bewegt er sich in der Folge immer weiter von dieser Sichtweise weg, bis es im Fazit schließlich – ganz im Sinne des hier thematisierten Irrtums – heißt:

EXMARaLDA [...] kann unter dieser Bedingung nicht mehr sein als ein Editor für das Schreiben in parallelen segmentierbaren Zeilen. (230)

EXMARaLDAs Metamorphose geht dabei nicht nur von einer "Sammlung von Konventionen und Werkzeugen" zu einem schlichten "Editor", sondern auch von eher (im texttechnologischen Sinne) inhaltlich/logischen Aspekten hin zu Aspekten der graphischen Präsentation. Da gerade diese Unterscheidung aber entscheidend ist für sämtliche moderneren Ansätze zur linguistischen Datenverarbeitung, sollen hier einige Stellen aus Schneiders Aufsatz exemplarisch besprochen werden.

⁵ Es scheint mir angemessen anzumerken, dass nicht alle am SFB verwendeten Transkriptionssysteme *gesprächsanalytische* Zielsetzungen verfolgen. Rein zahlenmäßig sind sogar die Transkriptionen zur (generativen) Spracherwerbsforschung weit überlegen. GAT kommt am SFB übrigens gar nicht zum Einsatz. Dass, wie Schneider sagt, EXMARaLDA dennoch mittlerweile "als allgemein verwendbares Transkriptionswerkzeug angeboten (und erwartet) wird" (229), entspricht eigentlich nicht der primären Zielsetzung des Projektes. Da EXMARaLDA aber in der Tat allgemein verwendbar sein sollte und konkurrierende, auf gesprächsanalytische Belange spezialisierte Projekte gar nicht existieren, sehe ich in dieser EXMARaLDA entgegengebrachten Erwartung aber auch kein ernsthaftes Problem.

⁶ Um es noch deutlicher zu sagen: die in einem binären, nicht offengelegten Format gespeicherten syncWriter-Daten sind ausschließlich für den syncWriter sinnvoll verwertbar; HIAT-DOS-Daten sind als "Text"-Daten zwar prinzipiell besser zugänglich, ihre konkrete Struktur macht eine Weiterbearbeitung aber ebenfalls schwer bis unmöglich. Die Entwickler dieser beiden Programme haben seit etwa zehn [!] Jahren keine Bemühungen erkennen lassen, an diesem Umstand etwas zu ändern.

den, an denen meines Erachtens unklar bleibt, ob von der Sprache EXMARaLDA oder "nur" dem Partitur-Editor die Rede ist.

Beispielsweise heißt es (218):

[...] sind EXMARaLDA-Transkripte aus Partiturzeilen (tiers) aufgebaut, die wiederum in Zellen zerlegt werden. [...] EXMARaLDA-Transkripte bestehen also aus Tabellen-Strukturen mit durchgängig definierten Zeilen und nicht durchgängig definierten Spalten. [In der zugehörigen Fußnote:] Die Strukturen anderer Partitur-editoren [sic] sind weniger tabellenförmig [...]

Die hier erwähnten Partiturzeilen, Tabellen, Zellen, Spalten und Zeilen sind Elemente graphischer Repräsentationen. Ein Gespräch hingegen ist ebenso wenig aus Partiturzeilen aufgebaut wie es Zellen, Spalten oder Zeilen enthält. Dort, wo EXMARaLDA die logische Struktur von Gesprächstranskriptionen kodiert, tauchen diese Elemente daher auch nicht auf. In der *basic-transcription* ist vielmehr von "events" (also: zeitbezogenen Ereignissen) die Rede, die *segmented-transcription* kennt die Bezeichnung "segment" (um sprachliche Einheiten von den rein zeitbezogenen "events" terminologisch abzugrenzen). Alle diese Einheiten werden über eine "timeline" zueinander in Bezug gesetzt. Dass der Partitur-Editor diese Elemente und ihre Beziehungen zueinander mit Hilfe einer Tabelle visualisiert, ist zwar richtig, es ist aber eben eine Eigenschaft des Editors und nicht von EXMARaLDA-Transkriptionen.⁷

Weiter unten im selben Abschnitt heißt es (218):

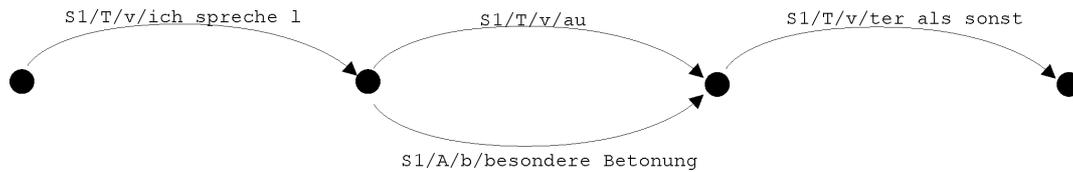
Partiturzeilen [...] sind lose über die Sprecherzugehörigkeit miteinander verbunden.

Was zwei "Partiturzeilen", also z.B. verbale und non-verbale Handlungslinien eines Sprechers verbindet, ist neben dem Bezug auf eine gemeinsame Zeitachse nichts weiter als eben diese Sprecherzugehörigkeit – die Verbindung ist daher sicherlich nicht als "lose" zu bezeichnen. Aus den weiteren Ausführungen, in denen Vergleiche zu den Sprecherblöcken in HIAT-DOS gezogen werden, wird deutlich, dass Schneider unter einer weniger losen Verbindung eine *räumliche* Nähe der betreffenden Elemente in der Präsentation zu verstehen scheint. Diese ist durch Umordnen der Spuren im Editor problemlos herzustellen, gehört aber mit Sicherheit nicht in die logische Repräsentation der Daten.

Noch undeutlicher wird die Unterscheidung zwischen den Strukturen, die die Sprache EXMARaLDA vorsieht, und denjenigen, die Eingabemaske und Ausgabemethoden des Partitur-Editors graphisch umsetzen, bei dem eigentlich zentralen Teil von Schneiders Aufsatz, dem Abschnitt "2.3. Strukturtypen für Annotate in EXMARaLDA." Obwohl im ersten Teil des Aufsatzes deutlich und richtig von "*Darstellungsmitteln* zum Annotieren" (Abschnitt 1.3.) gesprochen wird, werden logische Struktur und graphische Darstellung bei der exemplarischen Analyse von EXMARaLDA meines Erachtens nicht streng genug getrennt. Schneider zeigt dort – durchaus richtig –, dass die Ein- und Ausgabemethoden des Partitur-Editors einerseits einige der im ersten Teil klassifizierten Annotationsmethoden (z.B. An-

⁷ Ich wage übrigens zu behaupten, dass auch für den Partitur-Editor der Vergleich zu einer Tabelle bereits wesentlich weniger nahe liegt, wenn Gitternetzlinien ausgeblendet und eine einheitliche Hintergrundfarbe verwendet werden – der interessierte Leser möge das ausprobieren...

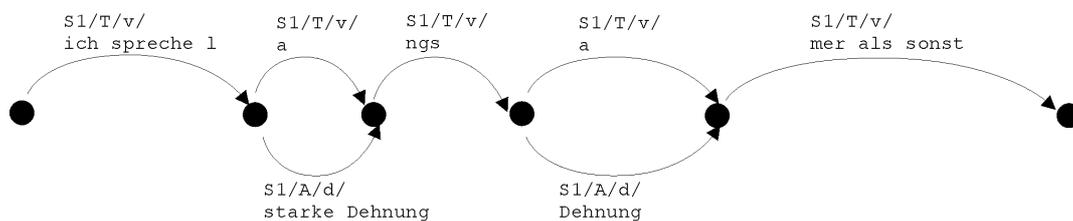
Logische Struktur:



Graphische Repräsentation:

Ich spreche la::ngsa:mer als sonst

Logische Struktur:



Die Kodierung dieser Strukturen ist in EXMARaLDA also ohne Weiteres möglich¹⁰ und ließe sich auch im Partitureditor vornehmen – ebenso wäre es möglich, aus dieser Struktur die entsprechenden graphischen Repräsentationen zu berechnen, der Editor leistet das allerdings z.Z. nicht. Die weiter unten stehende Anmerkung

Dass es sich beim Inhalt der zweiten Spur um die zur ersten Spur gehörenden Annotationen handelt, ist in den Datenstrukturen von EXMARaLDA nicht zwingend festgelegt. (221)

ist falsch. Wie in der obigen Darstellung berücksichtigt, sehen die Datenstrukturen von EXMARaLDA eine Typisierung der Spuren unter anderem nach T(ranskription) und A(nnotation) vor. Dass es sich beim Inhalt einer Spur um die zu einer anderen Spur gehörenden Annotation handelt, ist damit (in Kombination mit der Sprecherzuordnung) problemlos ersichtlich.¹¹

Zusammenfassend lässt sich sagen, dass Schneider eigentlich kritisiert, dass der *Partitur-Editor* nicht die gewünschten graphischen Repräsentationen aus den logischen Strukturen berechnen kann. Die Art und Weise, wie er diese Kritik formuliert, erweckt aber den Eindruck, dass die *Sprache* EXMARaLDA erst gar nicht für solche Strukturen konzipiert sei. Das ist falsch und bis zu einem gewissen Grad sicherlich auf den hier thematisierten Irrtum – nämlich der Gleichsetzung des Partitur-Editors mit den Möglichkeiten und Grenzen von EXMARaLDA – zurückzuführen.

Der zweite häufig anzutreffende Irrtum lautet:

¹⁰ Probleme ergeben sich erst an einer ganz anderen Stelle, nämlich dort, wo bei sich überlappenden Sprecherbeiträgen keine vollständige zeitliche Ordnung aller (sprachlichen) Elemente mehr möglich ist. Dieses Problem findet bei Schneider aber ebenso wenig Erwähnung wie der in Schmidt (2002a) ausführlich diskutierte zugehörige Lösungsansatz.

¹¹ Dass der "Zeilen-type" ein "funktionsloses Etikett" sei (221, Fußnote 65), ist also ebenfalls nicht richtig. Schneider begründet diese Behauptung auch nicht.

3.2. "XML ist die Lösung aller Probleme"

Seit XML vor etwas mehr als drei Jahren seinen Siegeszug um die Computerwelt angetreten hat, finden sich oftmals Lobpreisungen wie die folgende aus der Einführung von Ray (2001):

Mit XML kann so ziemlich jede Art von Information in einer Form gespeichert und organisiert werden, die auf Ihren Bedarf zugeschnitten ist. Als offener Standard ist XML weder an das Wohl und Wehe eines einzelnen Unternehmens gebunden, noch ist es mit einer bestimmten Software verknüpft. [...] Mit seiner klaren, einfachen Syntax und seinen eindeutigen Strukturen ist XML problemlos sowohl von Menschen als auch von Programmen zu lesen und zu analysieren. XML lässt sich mühelos mit Stylesheets kombinieren, um Dokumente im Format Ihrer Wahl zu erstellen. [...]

Auf den Punkt gebracht wird hier gesagt, dass XML eine für die Kodierung jeglicher Form von Daten gleichermaßen ideale Technologie darstellt. Schneider geht nicht ganz so weit, sagt aber immerhin (210):

Wenn man dann noch eine Strukturbeschreibung verwendet, die "selbsterklärend" und deren Syntax validierbar ist, kann man später mit einfachen Mitteln alle Informationen zu den Transkripten auslesen und sie in die gewünschte Form bringen. Eine Strukturbeschreibung mit [XML] ist heute wohl der beste Weg, obige Anforderungen "im Prinzip" zu erfüllen.

Im Fazit heißt es dann (230):

Einen Editor für Transkriptionen in einem XML-Format zu verwenden, aber ohne das [...] Markup an den eigentlich spannenden Stellen einzusetzen, wäre paradox. [...] Ein bisschen fortschrittlich werden und XML-Strukturen verwenden, aber ansonsten in den Annotationsstrukturen der Transkripte alles so zu lassen, wie es immer schon war, wäre inkonsequent.

Es ist mir auch nach mehrmaliger Lektüre des Artikels nicht ganz klar geworden, ob dies sich auf das Konzept von EXMARaLDA selbst bezieht, oder ob einfach vor einem inkonsequenten Gebrauch der potentiell in EXMARaLDA vorhandenen Möglichkeiten gewarnt wird. Implizit wird hier aber in jedem Falle suggeriert, dass es (mindestens) eine konsequente Art und Weise gäbe, Diskurstranskriptionen unter Ausnutzung aller Mittel von XML zu kodieren, und dass EXMARaLDA diese nicht zu nutzen wisse. Leider liegen die Dinge aber nicht so einfach. Um dies zu erläutern, muss ich etwas weiter ausholen.

Ein Großteil der im obigen Zitat von Ray genannten Vorzüge von XML, auf die auch Schneider verweist, verlieren ihre Gültigkeit, sobald man es mit Daten zu tun hat, deren primäre Struktur nicht der sogenannten OHCO-These genügt. Diese besagt, dass jedes Dokument eine "ordered hierarchy of content objects", also eine geordnete Hierarchie von Inhaltselementen, darstellt. Die OHCO-These ist bereits für viele Anwendungen auf Dokumente geschriebener Sprache leicht zu widerlegen, vgl. dazu z.B. Witt (2002). Für Transkriptionen, die ja Strukturen gesprochener Sprache beschreiben, ist ihre Ungültigkeit offensichtlich: Parallele (also eben nicht hierarchisch geordnete) Strukturen sind nicht nur ein Merkmal, das gesprochene von geschriebener Sprache unterscheidet, sie sind auch für die Repräsentationsformen vorhandener Transkriptionssysteme von ausschlaggebender Bedeutung – der oft thematisierte Vorteil der Partiturdarstellung besteht gerade darin,

dass sie die Darstellung paralleler Strukturen auf der obersten Ebene ihrer graphischen Organisation berücksichtigt. Weil die OHCO-These für Transkriptionen also keine Gültigkeit besitzt, gibt es auch keine "natürliche" Art und Weise, Transkriptionen als XML-Dokumente zu kodieren. Aus diesem Grunde kann bei einer XML-kodierten Transkription weder von einer "auf den Bedarf zugeschnittenen Datenorganisation", noch von "problemloser Lesbarkeit für den Menschen" oder einer "müheleisen Kombination mit Stylesheets zum Erstellen von Dokumenten" die Rede sein. All diese Attribute treffen nur auf solche XML-Dokumente zu, deren Datenstrukturen von vorneherein am ehesten durch Bäume beschrieben werden können. Der Vorteil des Einsatzes von XML zur Kodierung von Transkriptionen beschränkt sich demnach auf die Tatsache, dass es sich bei XML um einen offenen Standard handelt, der – im Vergleich zu beispielsweise SGML – in kürzester Zeit breiteste Anwendung gefunden hat und von dementsprechend vielen Programmen, Betriebssystemen und Entwicklungswerkzeugen unterstützt wird. Das ist mit Sicherheit ein nicht zu überschätzender Fortschritt,¹² der in jedem Falle genutzt werden sollte; alle Schwierigkeiten der Verarbeitung von Sprachdaten können damit aber nicht und schon gar nicht "mit einfachen Mitteln" gelöst werden.¹³

Der aktuelle Stand der diesbezüglichen Forschung geht dann auch über die bloße Anwendung von XML-Technologie hinaus. Vereinfacht ausgedrückt existieren zwei konkurrierende Ansätze, die beide die XML-eigenen Beschränkungen bezüglich nicht-hierarchischer Strukturen zu überwinden suchen – der Formalismus der Annotationsgraphen (Bird/Lieberman 2001), indem er statt von Baumstrukturen von gerichteten Graphen ausgeht, und das Konzept der Standoff-Annotation (Carletta et al. 2002), das auf periphere Technologien wie XPath und XLink setzt, um die Möglichkeiten von XML zu erweitern. Ersterer wird von Schneider beiläufig erwähnt, letzteres nur indirekt in einer Fußnote.

EXMARaLDA arbeitet mit dem Ansatz der Annotationsgraphen, der XML nur in seiner Eigenschaft als standardisiertem Speicherungsformat gebraucht, nicht aber in der Annahme, linguistische Annotation sei primär hierarchisch strukturiert. Dafür, dass bei EXMARaLDA in diesem Sinne und entgegen Schneiders Einschätzung konsequent vorgegangen wird, spricht meines Erachtens nicht zuletzt die Tatsache, dass vergleichbare Projekte wie TASX (Milde/Gut 2001), ELAN (MPI Nijmegen), ANVIL (Universität Saarbrücken) oder auch das altehrwürdige Praat für ihre Transkriptionssysteme bzw. -werkzeuge mit ganz ähnlichen, vom Prinzip her sogar identischen, Datenstrukturen arbeiten (vgl. dazu auch

¹² Er bedeutet nämlich, dass damit ein zuverlässiger und langfristiger Datenaustausch sichergestellt ist, ganz anders als es zum Beispiel für Word-Dokumente der Fall ist. Schneiders diesbezügliche Formulierung könnte unklarer kaum sein:

"Transkripte [mit dem Partitur-Editor] zu erstellen, garantiert deren Austauschbarkeit in etwa so zuverlässig – sogar längerfristiger –, wie den Austausch von Transkriptionen, die mit Word erstellt wurden" (230).

¹³ Eigentlich demonstriert Schneider das selbst. In Abschnitt 2.5. kritisiert er, dass beim Transkribieren nach zeitlichen Strukturen "Worte in verschiedene Zellen zerfallen" und dass diese erst beim Übergang in die *segmented-transcription*, die zusätzlich auch die sprachliche Struktur kodiert, wieder "zu ganzen Wort-Segmenten zusammengesetzt werden.", was er als "redundant" erachtet (224). Dies ist jedoch ein typisches Beispiel für die parallelen Strukturen, die in gesprochener Sprache auftauchen, und damit auch ein typischer Problemfall für die Kodierung dieser Strukturen in XML. Es mag bessere Lösungen dafür geben, als die in EXMARaLDA gewählte, Schneider hat aber keine solche zu bieten.

Schmidt i.V.). Diese Datenstrukturen mögen – wie Schneider wohl auch sagen möchte – nicht der Weisheit letzter Schluss sein, sie entsprechen aber zumindest der derzeit gängigsten Arbeitshypothese, die sich durchaus als praktikabel erwiesen hat. Schneiders Artikel erwähnt dies zwar indirekt (in Abschnitt 2.8.), wirklich über diese Arbeitshypothese hinausgehende Lösungsansätze bleibt er aber schuldig.

4. Softwareentwicklung für die Gesprächsforschung

Schneiders Aufsatz bleibt bei allem erkennbaren Praxisbezug eine eher theoretische Abhandlung. Er liefert Kriterien, an denen sich computergestützte Transkriptionssysteme messen lassen müssen und kritisiert den mit EXMARaLDA eingeschlagenen Weg. Zu einem oder sogar *dem* in der Praxis entscheidenden Punkt äußert er sich aber nur sehr vorsichtig: jedes noch so umfassende und theoretisch gut begründete Konzept, selbst wenn es im engagierten und andauernden Austausch zwischen Gesprächsforschern und Informatikern entwickelt wurde, ist wenig wert, wenn nicht auch die Mittel zu seiner Umsetzung und, was letztendlich genauso wichtig ist, die Mittel zu seiner Pflege und Anpassung an die sich rasant wandelnde Computerlandschaft sichergestellt sind. Dass für die EXMARaLDA-Entwicklung immerhin auf mehrere Jahre eine volle Mitarbeiterstelle zur Verfügung steht, ist wahrscheinlich ein in der deutschen Sprachwissenschaft bisher kaum gekannter Luxus.¹⁴ Im internationalen Vergleich oder gemessen an den Maßstäben einer professionellen Software-Entwicklung ist es dennoch schlicht unzureichend.¹⁵ Ob sich das innerhalb der deutschen Forschungslandschaft ändern ließe, kann ich nicht beurteilen. Ich bin mir aber sicher, dass sich der Arbeitsaufwand schon allein dann merklich reduzieren ließe, wenn Erfahrungen aus früheren, vergleichbaren Projekten, genutzt werden könnten. Das beste, weil irgendwie für den deutschsprachigen Raum spezifische Beispiel dafür ist der Partitur-Editor, der mindestens der sechste Versuch seiner Art ist.¹⁶ Bei der Entwicklung aller Vorgänger mussten immer wieder aufs Neue Benutzeranforderungen definiert, bestimmte vorwiegend technische Probleme (z.B. Umbruch nach Partiturflächen) gelöst, und ein Kompromiss zwischen Machbarem und Wünschenswertem gefunden werden. Zehn, fünf oder auch nur zwei Jahre später lassen sich diese Lösungs- und Entscheidungsfindungsprozesse aber so gut wie nicht mehr nachvollziehen, geschweige denn konkrete Lösungen wiederverwenden oder wenigstens Aufwandschätzungen anhand der vergangenen Erfahrungen vorneh-

¹⁴ Eine Ausnahme mag die langjährig gesicherte Arbeit des IDS an DIDA darstellen. Das erklärte Ziel, "die Bereitstellung eines Hilfsmittels, das die Verschriftlichung [...] von Sprachaufnahmen und damit den Aufbau von Korpora gesprochener Sprache unterstützt" (siehe Homepage), ist bislang allerdings nicht erreicht worden. Das IDS stellt meines Wissens keine öffentlich zugänglichen Transkriptionswerkzeuge zur Verfügung.

¹⁵ Zum Vergleich: die SourceForge-Webseite (<http://agtk.sourceforge.net/>) des US-amerikanischen AG-Toolkit-Projektes, das Software basierend auf dem Annotationsgraphenmodell entwickelt, listet vierzehn [sic] Entwickler auf.

¹⁶ Die fünf anderen mir bekannten sind: syncWriter, HIAT-DOS, der DIDA-Partitur-Editor, TranScript (Universität Halle) und PUGATE (Universität Koblenz). Interessanterweise scheint das Ausland hingegen an Partitur-Editoren relativ uninteressiert zu sein. Über einen US-amerikanischen Partitur-Editor im eigentlichen Sinne (d.h. über das, was gemeinhin als "Interlinear Text" bezeichnet wird, hinausgehend) ist mir z.B. nichts bekannt.

men. Über die Gründe dafür kann nur spekuliert werden; das grundsätzliche Problem scheint mir aber zu sein, dass die Entwicklung von Transkriptionswerkzeugen – bemerkenswerterweise ganz im Gegensatz zur Entwicklung von Transkriptionskonventionen – nie zu den eigentlichen *wissenschaftlichen* Aufgaben der Gesprächsforschung gezählt wurde und anderweitig gültige Praktiken der öffentlich zugänglichen Dokumentation somit keine Anwendung fanden. In diesem Zusammenhang möchte ich Schneider noch einmal ausdrücklich zustimmen, wenn er sagt (231):

Solange sich Gesprächsforscher nicht ebenso engagiert und detailliert an den DV-technischen Spezifikationen beteiligen, wie sie dies bei der Definition von Transkriptionskonventionen taten bzw. immer noch tun, wird es auch kein geeignetes Datenformat für Transkriptionen geben, das gleichzeitig in möglichst vielen Projekten mit unterschiedlichen Annotationsbedürfnissen einsetzbar ist. [...] Nur ein großer Interessentenkreis, der das Werkzeug in unterschiedlichen Projekten verwenden könnte, rechtfertigt den Einsatz der nicht unerheblichen Mittel, ein solch flexibles Werkzeug nach professionellen Maßstäben zu erstellen, kontinuierlich zu pflegen und weiterzuentwickeln.

Mit etwas optimistischerem Grundton möchte ich dem aber hinzufügen, dass es meines Erachtens verschiedene Anzeichen gibt, dass diesem Aufruf teilweise bereits nachgekommen wird. Beispielsweise finden sich in dieser Zeitschrift unter der Rubrik "Beiträge zur Forschungspraxis" bereits mehrere Aufsätze, die zumindest über derzeit vorhandene Möglichkeiten der computergestützten Transkription ausführlich informieren. Weiterhin war die Entwicklung von EXMARaLDA den letzten fünf Workshops der Funktionalen Pragmatik einen Beitrag wert, und auch auf anderen gesprächsanalytisch orientierten Workshops und Tagungen fanden "DV-technische" Aspekte des Transkribierens in letzter Zeit Eingang (vgl. z.B. mehrere Tagungsberichte in der letzten und aktuellen Ausgabe dieser Zeitschrift). Schließlich hat sich unter dem Terminus der Texttechnologie¹⁷ ein interdisziplinärer Forschungszweig entwickelt, der genau an der Schnittstelle zwischen Informatik und Linguistik ansetzt, die für den hier diskutierten Gegenstand relevant ist. Wenn diese Formen der Auseinandersetzung fortgesetzt und ausgeweitet werden, darf man in der Tat – im Sinne des nachfolgenden Zitates – auf weiteren Fortschritt bei der Entwicklung von Transkriptionswerkzeugen hoffen (228):

[Ein Teil der Transkribenten wird] auf einen Fortschritt bei der Entwicklung von Transkriptionseditoren hoffen – ob der EXMARaLDA-2 oder völlig anders heißt, ist egal.

Der Name, den dieser Fortschritt dann erhält, ist sicherlich wirklich zweitrangig. Nicht völlig gleichgültig erscheint mir aber, ob die Entwicklung wieder bei Null begonnen werden muss, oder ob in der bisherigen Arbeit an EXMARaLDA nicht doch weiterverwertbare Konzepte stecken. Konstruktive Kritik an diesen Konzepten wäre dann natürlich unverzichtbar.

¹⁷ Vgl. z.B. Lobin (1999) oder auch <http://www.text-technology.de>.

5. Literatur

- Bird, Steven / Liberman, Mark (2001): A formal framework for linguistic annotation. In: *Speech Communication* 33 (1,2), 23-60.
- Carletta, Jean / Isard, Amy / McKelvie, David (2000): *Linguistic Data Processing For Everyman*. Paper presented at the workshop on Web-Based Language Documentation and Description, Philadelphia, USA.
- Lobin, Henning (Hrsg.) (1999): *Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering*. Wiesbaden: Westdeutscher Verlag.
- Milde, Jan-Torsten / Gut, Ulrike (2001): The TASX-Environment: an XML-based corpus database for time aligned language data. In: Bird, Steven / Buneman, Peter / Liberman, Mark (Hrsg.) (2001): *Proceedings of the IRCS Workshop On Linguistic Databases*. Pennsylvania, 174-180.
- Ray, Erik (2001): *Einführung in XML*. (Deutsche Übersetzung von Olaf Brodacki). Köln: O'Reilly.
- Schmidt, Thomas (2002a): EXMARaLDA – ein System zur Diskurstranskription auf dem Computer. In: *Arbeiten zur Mehrsprachigkeit (Working Papers in Multilingualism), Serie B (34)*. Hamburg.
- Schmidt, Thomas (2002b): Gesprächstranskription auf dem Computer – das System EXMARaLDA. In: *Gesprächsforschung (Online-Zeitschrift zur verbalen Interaktion) 3*, 1-23.
- Schmidt, Thomas (i.V.): EXMARaLDA – ein System zur computergestützten Diskurstranskription. Erscheint in: Mehler, Alexander / Lobin, Henning (Hrsg.): *Werkzeuge zur automatischen Analyse und Verarbeitung von Texten: Formate, Tools, Software-Systeme*.
- Schneider, Wolfgang (2002): Annotationsstrukturen in Transkripten – DV-technische Strukturanforderungen für Annotate exemplifiziert an EXMARaLDA. In: *Gesprächsforschung (Online-Zeitschrift zur verbalen Interaktion) 3*.
- Witt, Andreas (2002): *Multiple Informationsstrukturierung mit Auszeichnungssprachen. XML-basierte Methoden und deren Nutzen für die Sprachtechnologie*. Dissertation, Universität Bielefeld.

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

Veröffentlicht am 23.2.2003

© Copyright by GESPRÄCHSFORSCHUNG. Alle Rechte vorbehalten.