

## **Transkriptionskonventionen für die computergestützte gesprächsanalytische Transkription**

**Thomas Schmidt**

### **Einleitung**

Die [GAT-]Konventionen richten sich an Transkribenten, nicht an Computer. (Selting et al. 1998)

Ein geeignetes Transkriptionsprogramm zu finden, ist schwierig. Die erhältlichen sind eher unhandlich; die meisten können zudem nicht für die Transkription mit GAT verwendet werden. (Linke et al. 2004)

Gesprächsanalytische Transkription wird heute fast ausnahmslos mit Hilfe des Computers bewerkstelligt. Die methodischen Grundlagen der Transkription, insbesondere die auch heute gebräuchlichen Transkriptionsrichtlinien und –konventionen, haben jedoch ihren Ursprung in einer Zeit, in der das Transkribieren eine mit Hilfe von Bleistift (oder Schreibmaschine) und Papier ausgeübte Tätigkeit war. Zwar wurden die Konventionen stellenweise den neuen medialen Rahmenbedingungen angepasst, dies geschah aber eher in der Absicht, die etablierten Verfahren möglichst unverändert im neuen, digitalen Medium nutzbar zu machen, als mit dem offensiven Vorsatz, sie für die neuen Möglichkeiten ebendieses Mediums grundlegend zu überarbeiten und zu optimieren. Das daraus resultierende Dilemma bringen die beiden einleitenden Zitate auf den Punkt: auf der einen Seite stehen Transkriptionskonventionen, in denen – teilweise bewusst – auf eine Auseinandersetzung mit modernen computergestützten Verfahren der Transkription verzichtet wird; auf der anderen Seite wird beklagt, dass existierende Transkriptionssoftware die Arbeit mit den ebendiesen Konventionen nur unzureichend unterstütze.

Der vorliegende Beitrag macht einige Vorschläge, wie dieses Dilemma durch eine Überarbeitung vorhandener Transkriptionskonventionen aufgelöst werden könnte. Er stützt sich wesentlich auf die Erfahrungen, die bei der Entwicklung von EXMARaLDA (Schmidt/Wörner 2005) mit den Konventionen der Halbinterpretativen Arbeitstranskriptionen (HIAT, Ehlich/Rehbein 1976 und 1979, Rehbein et al. 1993) gemacht wurden. Auch dabei hat sich nämlich gezeigt, dass die bestehenden Versionen der Konventionen viele Schritte der halb- oder vollautomatischen Bearbeitung von digitalen Transkriptionen (z.B. automatische Segmentierung, Suche auf Transkriptionskorpora, Austausch von Transkriptionen zwischen verschiedenen Anwendungen) schwierig oder unmöglich machen, weil sie Grundprinzipien der rechnergestützten Datenverarbeitung nur unzureichend berücksichtigen. Es wurde daher eine neue Fassung der Konventionen erarbeitet (Rehbein et al. 2004), die zumindest die wichtigsten diesbezüglichen Mängel behebt. Auf diese Überarbeitung der HIAT-Konventionen werde ich im Verlauf dieses Beitrags wiederholt exemplarisch verweisen.

Eine detaillierte Analyse von HIAT und zwei weiteren für die Gesprächsforschung wichtigen Transkriptionskonventionen (GAT, Selting et al. 1998 und

DIDA,<sup>1</sup> Klein/Schütte 2001) im Hinblick auf ihre Eignung für computergestützte Verfahren (Schmidt 2005) hat gezeigt, dass sich die Probleme über verschiedene Systeme hinweg gleichen. Insofern sollte vieles, was in diesem Beitrag vornehmlich anhand von HIAT exemplifiziert wird, in ähnlicher Weise für GAT, DIDA und andere, aus der konversations- oder diskursanalytischen Forschungstradition entstandene Konventionen<sup>2</sup> gelten.

Dieser Beitrag gliedert sich in fünf Abschnitte, in denen jeweils ein Bereich diskutiert wird, der für die Konzeption und Implementierung von Instrumenten zur computergestützten Transkription relevant ist.

## 1. Zeichenebene

Eine der grundlegendsten (und gleichzeitig am einfachsten zu erfüllenden) Voraussetzungen für eine computergestützte Verarbeitbarkeit von Transkriptionen ist eine eindeutige Definition der zu verwendenden Transkriptionszeichen. Diese Definition sollte zum einen über eine Zuordnung von zu verwendenden graphischen Zeichen zu einer Position in der Unicode-Zeichentabelle<sup>3</sup> erfolgen. In aller Regel, etwa für die Symbole des lateinischen Alphabets, ist dies eine triviale Aufgabe. Bei manchen Zeichen, insbesondere solchen, die typischerweise benutzt werden, um speziell Eigenschaften gesprochener Sprache zu repräsentieren, gibt es jedoch Mehrdeutigkeiten zu klären. Hierfür einige Beispiele aus HIAT:

- Tonbewegungen werden in HIAT über ein diakritisches Zeichen auf dem Vokal der betreffenden Silbe repräsentiert (30f.).<sup>4</sup> Viele dieser Vokal-Diakritikon-Kombinationen existieren jedoch auch als Einzelzeichen ("Ligaturen") in der Unicodetabelle, beispielsweise die u.a. im Französischen gebräuchliche Kombination aus *e* und Akut (wie in *été*). In diesem Falle wurde für die überarbeitete Version der Konventionen festgelegt, dass zum Kennzeichnen von Tonbewegungen immer eine Kombination aus dem betreffenden Zeichen (etwa U+0065<sup>5</sup> für *e*) und einem "*combining diacritical mark*" (etwa U+0301 für den Akut) zu verwenden ist, und nicht die Ligatur (etwa U+00E9 für *é*). Zu begründen ist diese Entscheidung dadurch, dass die betreffenden Zeichen sich auf unterschiedliche (segmentale bzw. nicht-segmentale) Eigenschaften des Transkribierten beziehen und es daher sinnvoller erscheint, sie auch auf der Zeichenebene als getrennte Einheiten zu kodieren.

<sup>1</sup> Auch für DIDA liegt mittlerweile eine überarbeitete Fassung (Schütte 2004) für die Arbeit mit EXMARaLDA vor.

<sup>2</sup> Zurzeit arbeiten wir an einer Konvertierung des Santa Barbara Corpus of Spoken American English (DuBois et al. 2000-2005) nach XML. Auch mit dem dabei verwendeten Transkriptionssystem der Discourse Transcription (DuBois 1991) stellen sich vergleichbare Probleme.

<sup>3</sup> Unicode ist eine mit dem ISO-Standard 10646 koordinierte Initiative, die Festlegungen für eine vereinheitlichte Repräsentation von Schriftzeichen auf dem Computer trifft. Unicode ist – zumindest was die für die Transkriptionen relevanten Bereiche anbelangt – ein ausgereifter, stabiler und von den allermeisten modernen Anwendungen unterstützter Standard. Ihn auf dieser Ebene als Definitionsbasis zu setzen, dürfte daher kaum kontrovers sein.

<sup>4</sup> Hier und im Folgenden beziehen sich Seitenangabe ohne Literaturreferenz auf Rehbein et al. 2004.

<sup>5</sup> Dies ist die üblichste Notation für Unicode-Zeichen. Sie gibt die Position des Zeichens in der Unicode-Tabelle in hexadezimaler Schreibweise an. 'U+0065' bezeichnet also das Zeichen an Position 101 (dezimal) der Unicode-Tabelle.

- Der Abbruch einer Äußerung wird in HIAT durch den dreifachen Ellipsenpunkt (*Aber ich...*) kenntlich gemacht (22ff.). Auch hier besteht eine Alternative zwischen einer dreifachen Verwendung des Punktzeichens (U+002E) oder der in Unicode vorhandenen Verbindung dieser Zeichen zu einem einzigen (U+2026). In diesem Falle wurde festgelegt, die Verwendung des verbundenen Einzelzeichens zu fordern. Auf diese Weise wird das für andere Äußerungstypen gültige Prinzip – "(genau) *ein* abschließendes Zeichen charakterisiert den Äußerungstyp" – gewahrt und potentielle Verwechslungen mit dem Punkt, der in HIAT eine nicht abgebrochene deklarative Äußerung abschließt, vermieden.
- Einige Phänomene gesprochener Sprache werden in HIAT durch Zeichen repräsentiert, die im schriftsprachlichen Gebrauch eher unüblich sind. Dazu gehört der Ligaturbogen zum Kennzeichnen eines schnellen Anschlusses (50) oder der Punkt ("Bullet") zum Kennzeichnen einer Pause (34f.). In diesen Fällen musste für die Überarbeitung der Konventionen einfach ein geeignetes Zeichen aus der Unicode-Tabelle (U+203F bzw. U+2022) gesucht und verbindlich festgelegt werden.

Zum anderen bedeutet die Eindeutigkeit von Transkriptionszeichen auch, dass nicht identische Zeichen verwendet werden, um verschiedenartige Phänomene zu bezeichnen, und dass umgekehrt nicht verschiedene Zeichen verwendet werden, um ein- und dasselbe Phänomen zu beschreiben. Diese Anforderung ist im Prinzip auch unabhängig von einer computergestützten Verarbeitung von Transkriptionen sinnvoll und wird von den bestehenden Systemen auch weitestgehend eingehalten. Vereinzelt existieren jedoch Doppelzuweisungen, die ein menschlicher Benutzer eines Transkripts anhand ihres Kontextes problemlos auflösen kann, die für die computergestützte Verarbeitung aber problematisch sind. In HIAT war die Verwendung des Doppelpunkts, der nach Rehbein et al. (1993) einmal zum Kennzeichnen einer Vokallängung (*a:ber*), einmal zum Kennzeichnen einer Ankündigung (*er sagt: "Hallo"*) verwendet werden konnte, ein solcher Fall. In den überarbeiteten Konventionen wurde die Verwendung des Doppelpunkts auf letztere Funktion beschränkt und für die Vokallängung ein alternatives Verfahren (Verdoppelung des Zeichens: *aaber*) festgelegt.

## 2. Einheiten der Transkription über der Zeichenebene

Ebenso wichtig für die computergestützte Verarbeitbarkeit von Transkriptionen ist ein konsistentes und eindeutig definiertes System von Einheiten der Transkription oberhalb der Zeichenebene. Alle bestehenden Systeme definieren Kategorien für solche Einheiten (etwa: Wort, Pause, Äußerung etc.) und machen von ihnen auf vielfältige Weise in der theoretischen Fundierung sowie der Darstellung und Analyse von Transkriptionen Gebrauch.<sup>6</sup> Für den Aspekt der Computerunterstützung sind dabei folgende Punkte wichtig:

---

<sup>6</sup> Für eine ausführliche Diskussion, die diesbezügliche Eigenschaften von HIAT, GAT, DIDA im Detail analysiert und vergleicht, sei an dieser Stelle auf Schmidt (2005:113ff.) verwiesen.

1. die Einheiten müssen eindeutige Namen tragen
2. die Einheiten sollten untereinander strukturiert sein, etwa indem sie verschiedenen – und im Sinne der theoretischen Fundierung der Transkriptionskonvention sinnvollen – Ebenen zugeordnet werden
3. die Einheiten sollten systematisch kontrastiv sein (vgl. den Begriff der "Systematic Contrastiveness" bei Edwards 1992b:436), d.h. innerhalb einer Ebene des Systems sollten Einheiten einander nicht überschneiden.
4. die Einheiten auf jeweils einer Ebene des Systems sollten die Transkription vollständig ("exhaustiv") erfassen, d.h. es sollte nicht vorkommen, dass gewisse Transkriptionszeichen auf einer Ebene unklassifiziert bleiben
5. die möglichen Beziehungen zwischen zwei Einheiten sollten eindeutig definiert sein

Idealerweise würde also jede Transkriptionskonvention ein hierarchisches<sup>7</sup> System mit einer festen Anzahl von Ebenen definieren, das alle Transkriptionszeichen erschöpfend und eindeutig erfasst. In bestehenden Systemen ist dies bereits angelegt, so etwa für HIAT und GAT:

	<b>HIAT</b>	<b>GAT</b>
<b>0</b>	Diskurs	Gespräch
<b>1</b>	Turn	Turn
<b>2</b>	Äußerung	Phrasierungseinheit
<b>3</b>	Wort, Pause, nicht-phonologische Einheit	Wort, Pause, non-verbale Handlung
<b>4</b>	Silbe	Silbe

Studiert man die Konventionen im Detail, treten jedoch einige Unstimmigkeiten zutage, die sich als problematisch erweisen, wenn auf ihrer Grundlage ein System zur computergestützten Transkription konstruiert werden soll.

So werden beispielsweise in den GAT-Konventionen unter der Überschrift "Sequenzielle Struktur" (S.97f.) die Begriffe *Turn*, *Sprecherbeitrag*, *Intonationseinheit* und *prosodische Einheit* verwendet, um teilweise verschiedene, teilweise gleiche Einheiten der Transkription zu bezeichnen und zu erläutern, wie diese sich zur graphischen Einheit der *Zeile* bzw. *Transkriptzeile* verhalten. Später (S.100) kommt der Begriff *Phrasierungseinheit* hinzu und wird seinerseits zur Einheit der *Zeile* in Bezug gesetzt.

Zunächst gilt hier also nicht, dass Einheiten eindeutige Namen tragen. Zwar wird einem mit der Materie vertrauten Leser klar sein, dass *Turn* und *Sprecherbeitrag* einerseits und *Intonationseinheit*, *prosodische Einheit* und *Phrasierungseinheit* andererseits hier synonym gebraucht werden. Bei der Konstruktion von formalen Datenmodellen für die computergestützte Transkription ist eine solche terminologische Vielfalt aber unbedingt zu vermeiden, und die Konventionen sollten entsprechend *einen* bevorzugten Terminus für jede Einheit als solchen ausweisen.

<sup>7</sup> Streng genommen muss ein solches System eigentlich nicht zwangsläufig ein hierarchisches sein. Der Hierarchiegedanke ist aber in allen Formen der Sprachbeschreibung und -analyse derart dominant, dass es schwer vorstellbar ist, dass ein intuitiv nutzbares Transkriptionssystem nicht von ihm Gebrauch macht.

Weiterhin lässt sich aus den Konventionen nicht eindeutig ableiten, wie genau sich die Einheit der *Transkriptzeile* zu den Einheiten *Turn* und *Phrasierungseinheit* verhält. Für den Beginn einer neuen Zeile werden mindestens die folgenden Kriterien genannt: Beginn eines Turns (S.97), Beginn einer eindeutig zu identifizierenden Phrasierungseinheit (S.100), Neueinsetzen nach Simultansprechen (S.97), nicht eindeutig zuzuordnende Pause (S.98). Dabei bleibt unklar, ob das Neueinsetzen nach Simultansprechen (und damit ein Kriterium für den Beginn einer neuen Zeile) gleichbedeutend mit dem Beginn eines neuen Turns oder einer neuen Phrasierungseinheit ist. Gleichzeitig macht ein Beispiel (S.101) deutlich, dass eine Transkriptzeile auch eine Kombination von einer Phrasierungseinheit und einem nachfolgenden "unakzentuierten Anhängsel" (dessen Status als eigenständige Phrasierungseinheit laut den Autoren "umstritten" ist) beinhalten kann.

Implizit wird so durch die Zeile eine Einheit der Transkription definiert, die weder eindeutig benannt noch in eindeutiger Weise zu den anderen Einheiten in Bezug zu setzen ist. Zwar heißt es in einer Fußnote (S.120): "[...] garantiert dieses Verfahren [der Zeilenstrukturierung], dass Zeilenangaben sich i.a. auf eine sinnvolle Einheit beziehen [...]". Das Problem liegt hierbei aber im einschränkenden "i.a." – für eine computergestützte Verarbeitbarkeit ist es entscheidend, dass die hier behandelten Begriffe und Zusammenhänge entweder ohne Ausnahme gelten (also etwa: "eine Zeile entspricht einer Phrasierungseinheit / der Einheit XY"), oder dass mögliche Ausnahmen explizit benannt werden (also etwa: "eine Zeile entspricht einer Phrasierungseinheit, einer nicht zuzuordnenden Pause oder einer Phrasierungseinheit mit unakzentuiertem Anhängsel").

Ähnliche Probleme ergaben sich bei der Überarbeitung der HIAT-Konventionen bzgl. der Begriffe *Turn*, (*Sprecher-*)*Beitrag*, *Äußerung* und *Segment*. Hier entstanden zusätzliche Schwierigkeiten dadurch, dass der Begriff *Turn* innerhalb der Transkriptionskonventionen in einer "naiven" Lesart gebraucht wurde (vgl. Rehbein et al. 1993:34), die einer funktional-pragmatisch elaborierten Definition des gleichen Begriffes nicht gerecht wird: während erstere sich in etwa umschreiben lässt als "alles was ein Sprecher tut zwischen zwei Abschnitten, in denen er nichts tut", beinhaltet letztere theoretisch komplexe Konstruktionen wie ko-konstruierte Äußerungen, die zwar am Transkript zu rekonstruieren sind, bei der Transkription selbst aber keine relevante Einheit darstellen.<sup>8</sup> Als nun für die Überarbeitung der Konventionen nach einer präzisen Definition des Begriffes *Turn* (so wie er in den Konventionen gebraucht wird) gesucht wurde, erwies sich die Konkurrenz zwischen den beiden Lesarten als Hindernis. Es wurde daher beschlossen, in den Konventionen ganz auf den Begriff *Turn* zu verzichten und stattdessen von *Segmentketten* zu sprechen (vgl. Rehbein et al. 2004:18).

Zahlreiche weitere solcher Beispiele könnten hier ausgeführt werden, etwa die nirgendwo explizit definierte, aber für fast jeden Schritt der computergestützten

---

<sup>8</sup> In den GAT-Konventionen ist dies nicht grundsätzlich anders: dort wird mit "Turn" das bezeichnet, was im Transkript als eine Folge von Transkriptzeilen mit *einem* einleitenden Sprecherkürzel erscheint. Gleichzeitig macht Selting (2001:1061) deutlich, dass die Gesprächsanalyse eigentlich mit einer wesentlich komplexeren Definition des Turnbegriffes operiert: "Scheinbar ist ein Redebeitrag einfach zu definieren [als] die Rede eines Gesprächsteilnehmers zwischen Sprecherwechseln. [...] [Diese Definition] muss ersetzt werden durch eine Definition, in der der Turnbegriff in Zusammenhang gebracht wird mit den interaktiven Verfahren der Projektion und Herstellung von Turns durch Sprecher und Rezipienten in bestimmten sequentiellen Kontexten."

Transkription unabdingbare Einheit des *Wortes* oder das unklare Verhältnis zwischen den Einheiten *Wort* und *Silbe*. Sie alle verstoßen gegen eine oder mehrere der oben aufgeführten Anforderungen, und ihnen allen ist gemein, dass sie aus einer gewissen Nachlässigkeit beim Konstruieren und Beschreiben des terminologischen Systems für die Transkription erwachsen.<sup>9</sup> Beim Transkribieren mit Bleistift und Papier mag diese Nachlässigkeit sich entweder nicht bemerkbar machen oder als unerheblich erachtet werden. Für die computergestützte Transkription ist ein konsistent benanntes und klar organisiertes System von Transkriptionseinheiten jedoch ein dringendes Desiderat.

### 3. Inhalt vs. Darstellung

Beim Transkribieren auf Papier muss der Transkribent in dem Moment, in dem er Einheiten der Transkription identifiziert und beschreibt, auch direkt entscheiden, ob, wie und wo er diese auf seinem Medium darstellen möchte. Entsprechend sind in den herkömmlichen Transkriptionskonventionen Angaben zur Inhalts- und zur Darstellungsseite der Transkription miteinander vermischt. So werden beispielsweise in den älteren Fassungen der HIAT-Konventionen alle Transkriptionsanweisungen unmittelbar auf das "Notieren" in der graphischen Form der Partitur bezogen.

Dabei ist es unbestritten, dass die technischen Beschränkungen des Papiers und das Bestreben, Transkriptionen gut lesbar zu halten, den Transkribenden zu einer Reihe von Kompromissen zwingen: zum einen muss das mehrdimensionale Geschehen im Gespräch auf einer begrenzten zweidimensionalen Fläche wiedergegeben werden; zum anderen soll der Leser eines Transkripts zwar möglichst viele Informationen präsentiert bekommen, gleichzeitig aber nicht durch eine zu große Anzahl verschiedener Zeichen belastet werden. Deppermann (1999:41) sagt dazu:

An Transkriptionssysteme sind prinzipiell widersprüchliche Anforderungen gestellt: In praktischer Hinsicht sind einfache Lesbarkeit [...] und unproblematische Realisierbarkeit in gängigen Textverarbeitungssystemen wünschenswert, aus gegenstandsbezogen-theoretischen Gründen wünscht man sich dagegen Umfassendheit, Präzision und Repräsentation formbezogener Parameter, [...].

Es ist der für die Praxis vielleicht wichtigste Unterschied zwischen papier- und computerbasierter Transkription, dass bei letzterer der Widerspruch zwischen diesen Anforderungen nicht oder zumindest nur in sehr viel geringerem Maße bestehen muss. Das digitale Medium erlaubt es, Daten unabhängig von einer bestimmten Präsentationsform aufzubewahren und zu bearbeiten, und für ein einziges Datum im Prinzip beliebig viele Präsentationsformen automatisch zu erstellen. Dass dieses Grundprinzip der Trennung von Inhalt und Darstellung entscheidend ist für fast jede Form der rechnergestützten Verarbeitung sprachlicher Daten, ist an anderer Stelle ausführlich dargestellt (vgl. z.B. Witt 2002, Johansson 1995, Schmidt 2002) und soll hier nicht noch einmal wiederholt werden. Stattdessen seien einige konkrete Beispiele gegeben, wie bestehende gesprächsanalytische Transkripti-

---

<sup>9</sup> Hausendorf (2001:972) spricht davon, dass methodische Prinzipien in der Gesprächsforschung oft "eher beiläufig skizziert" werden. Nach den hier dargelegten Beobachtungen gilt das offenbar auch für die Transkriptionskonventionen.

onskonventionen zu ändern wären, damit sie von der Anwendung dieses Prinzips profitieren können.

Zunächst sollte es ein grundsätzliches Anliegen der Konventionen sein, systematisch und deutlich zu trennen zwischen Termini, die (inhaltliche) transkribierte Einheiten des Gesprächs bezeichnen und solchen, die (graphische) Elemente der Transkription benennen. Beispielsweise finden sich in Rehbein et al. (1993:59f.) die folgenden Unterabschnitte zum Thema "Interpunktion":

4.2.2. Ligatur: Eine Ligatur liegt vor, wenn eine Zäsur (notiert durch ein Grenzzeichen) [...] "übersprochen" wird [...] Man notiert nach dem betreffenden Satzzeichen zuerst eine Leertaste, dann [...] das Ligaturzeichen (= Bogen am unteren Zeilenrand nach vorn) [...] 4.2.3. Gedankenstrich: Der Gedankenstrich schließt eine Einfügung (Parenthese, Einschub) ein.

Dabei wird einmal das Phänomen ("Ligatur") in der Überschrift benannt und im nachfolgenden Text das zu seiner Notation zu verwendende Zeichen ("Bogen") spezifiziert. Im nachfolgenden Abschnitt wird dann umgekehrt das Transkriptionszeichen ("Gedankenstrich") in der Überschrift aufgeführt, während der zugehörige Text das Phänomen ("Parenthese") benennt. Solche Inkonsistenzen sind in der Überarbeitung der Konventionen behoben.

Weiterhin sollten Konventionen es vermeiden, gewisse sprachliche Phänomene nur aus der graphischen Darstellung der Transkription ableitbar zu machen, ohne sie auf der Inhaltsseite durch entsprechende Zeichen zu repräsentieren. Z.B. vermerken Rehbein et al. (1993:81f.) zur Transkription von "ungefüllte[n] Unterbrechungen":

Wir notieren [die Unterbrechung einer Äußerung] nicht mit einem besonderen Zeichen, sondern synchronisieren die unterbrochene Äußerung [...] an die parallel laufenden gefüllten Spuren [...]. Auf diese Weise erzeugen wir [...] relativ zur Dauer des Innehaltens eine Lücke bzw. mehrere Lücken. Die Unterbrechung wird nicht durch Pausenpunkte notiert.

Eine resultierende Partitur sieht nach dieser Anweisung etwa folgendermaßen aus:

X [v]	Das ist	halt so.
Y [v]		Mmh

Dass Sprecher X an der betreffenden Stelle eine äußerungsinterne Pause macht, ist aber nur an dieser Darstellung abzulesen. In der Transkriptionsdatei selbst (oder ggf. auch bei anderen Darstellungsformen) ist diese Information nicht vorhanden, denn sie ergibt sich nur *implizit* aus dem Zusammenspiel von Inhalt und Darstellung. In den überarbeiteten Konventionen wird daher für solche Unterbrechungen ein explizites Transkribieren der sich ergebenden Pause gefordert.

In ähnlicher Weise problematisch ist die in verschiedenen Systemen praktizierte Methode, die zeitliche Ausdehnung eines (meist non-verbalen) Phänomens mit Hilfe einer Folge von Füllzeichen (z.B. Bindestrich) zu beschreiben, wobei sich deren Anzahl danach richtet, wie viel Raum die graphische Repräsentation eines zeitgleichen (meist verbalen) Phänomens einnimmt. So würde nach Rehbein et al. (1993) etwa ein mimisches Phänomen folgendermaßen in einer HIAT-Partitur beschrieben:

<b>X [v]</b>	Na ja... Das ist halt so, oder? Oder nicht?
<b>X [nv]</b>	°--- Augenklimpern ---°

Bei einer Trennung von Inhalts- und Darstellungsaspekt gehört jedoch nur die Beschreibung "Augenklimpern" sowie Information über deren Anfangs- und Endpunkt (zeitgleich mit Beginn des Wortes "Das" bzw. mit dem Ende des Wortes "oder") auf die Inhaltsseite. Die verwendeten Füllzeichen sind hingegen auf der Inhaltsseite nicht sinnvoll interpretierbar – ihre Anzahl ergibt sich (in nicht systematisch vorhersagbarer Weise) aus Parametern der Darstellung wie den in den Spuren verwendeten Schriftsätzen und -größen, die ihrerseits gerade nicht auf die Inhaltsseite der Transkription gehören. In den überarbeiteten Konventionen wird daher auch diese Vermischung von Inhalt und Darstellung beseitigt.

Es ist also wichtig zu vermeiden, dass gewisse Informationen nur aus einer graphischen Darstellung heraus, und nicht aus dem Inhalt der Transkriptionsdatei selbst sinnvoll interpretierbar sind. Umgekehrt kann es aber durchaus sinnvoll sein, auf der Inhaltsseite Informationen festzuhalten, die bei gewissen Formen der graphischen Repräsentation weggelassen oder stark verdichtet dargestellt werden.

Eine solche Information ist die Angabe von absoluten Zeitwerten, die Stellen in der Transkription mit den zugehörigen Stellen in der Audio- oder Videoaufnahme verbinden. Während diese Angaben äußerst nützlich sein können, um beim Transkribieren oder bei der Analyse die Navigation in Transkription und Aufnahme mit Hilfe einer geeigneten Software zu erleichtern, sind sie beim Lesen eines Transkripts normalerweise eher hinderlich. In ähnlicher Weise mögen andere Arten von Information sich je nach Untersuchungszweck als mehr oder weniger darstellungswürdig erweisen – beispielsweise kann es für eine Wortschatzanalyse hilfreich sein, prosodische Informationen weitestgehend aus dem Transkript auszublenden, während es für eine Intonationsanalyse sinnvoll ist, die gleichen Informationen möglichst deutlich hervorzuheben.

Schließlich lässt sich auch die Opposition zwischen Partitur- und Zeilennotation, die in der Gesprächsforschung verschiedentlich als theoretisch fundamental behandelt wird (vgl. vor allem Rehbein 2001 und Selting 2001), weitestgehend auf das Prinzip "gleicher Inhalt, verschiedene Darstellungen" reduzieren. Wie beispielsweise das EXMARaLDA-System demonstriert, macht es eine inhaltsbezogene Kodierung von Transkriptionsdaten durchaus möglich, beide Darstellungsformen (und ggf. auch gänzlich anders organisierte Transkriptionslayouts) nebeneinander zu benutzen, ohne dabei die etablierten Transkriptionsgewohnheiten grundlegend zu ändern.<sup>10</sup>

Spätestens wenn man daran interessiert ist, die in der Transkription festgehaltenen Informationen in der Darstellung derart flexibel handhaben zu können, führt kein Weg mehr daran vorbei, den Transkriptionskonventionen eine formale Defi-

<sup>10</sup> An dieser Stelle sei angemerkt, dass – allen theoretischen Bedenken zum Trotz – sowohl HIAT- als auch GAT-orientierte Forscher sich nach meinem Eindruck in der Praxis ohnehin schon beider Layoutformen bedienen, je nachdem, welche von beiden sich für den speziellen Analysezweck als geeigneter erweist. Indem sich die Transkriptionskonventionen dennoch einseitig auf eine Layoutform festlegen, hinken sie also in diesem Punkt unabhängig von Erwägungen der Computerunterstützung bereits der gängigen Praxis hinterher.



tion eines inhaltsorientierten Datenformats an die Seite zu stellen.<sup>11</sup> Wird eine solche Definition sorgfältig vorgenommen, gehen damit die Möglichkeiten der computergestützten Verarbeitung von Transkriptionen sogar wesentlich über die hier angedeuteten hinaus.

#### 4. Transkription als Prozess

Es wird bereits in den bestehenden gesprächsanalytischen Transkriptionskonventionen deutlich festgestellt, dass das Transkribieren ein Prozess ist, in dem das resultierende Produkt *schrittweise* ausgebaut und verfeinert wird. HIAT trägt dem bereits in seinem Namen *Halbinterpretative Arbeitstranskriptionen* Rechnung, GAT in der Unterscheidung zwischen Basis- und Feintranskript und dem sog. Zwiebelprinzip (Selting et al. 1998:92):

Ausbaubarkeit und Verfeinerbarkeit der Notation ("Zwiebelprinzip"): Ein Transkript einer bestimmten Detailliertheitsstufe soll ohne Revision der weniger differenzierten Version ausbaubar und verfeinerbar sein.

Da jedoch, wie bereits oben beschrieben, beim Transkribieren auf Papier der Prozess des Transkribierens gleichbedeutend ist mit dem Erstellen eines weitestgehend statischen Endprodukts, sind die bestehenden Konventionen letzten Endes eher produkt- als prozessorientierte Anweisungen an den Transkribierenden. Werden hingegen Inhalt und Darstellung der Transkription in oben angedeuteter Weise konsequent getrennt, kann die Prozesshaftigkeit des Transkribierens auch in den Konventionen stärker berücksichtigt werden. Im Zusammenspiel mit geeigneten Softwarewerkzeugen kann dies die Effizienz und Präzision des Transkribierens wesentlich verbessern.

Man betrachte als Beispiel etwa die prosodisch detaillierte Transkription eines fremdsprachigen Gesprächs nach HIAT. Diese wird mindestens die folgenden Schritte beinhalten:

1. Transkription von geäußerten Wörtern (und von Pausen, paralinguistischen Merkmalen etc.)
2. Präzise zeitliche Ordnung der Beiträge verschiedener Sprecher (insb. bei Überlappungen)
3. Annotation prosodischer Eigenschaften (z.B. Akzente, Modulation)
4. Festlegen von Segmentgrenzen (d.h. Äußerungsgrenzen nach HIAT)
5. Hinzufügen von deutschen Übersetzungen für fremdsprachige Äußerungen

Diese Schritte bauen aufeinander auf (prosodische Eigenschaften werden an transkribierten Wörtern annotiert, Segmentgrenzen werden von prosodischen Eigenschaften mitbestimmt, die Übersetzung erfolgt auf der Basis der festgelegten Segmentgrenzen, etc.) und werden daher, auch wenn sie sich stellenweise wechselseitig beeinflussen mögen, in der Regel nacheinander ausgeführt. Dabei unter-

---

<sup>11</sup> In der Praxis erfolgt die Definition eines solchen Formats heute in aller Regel auf der Grundlage von XML, das den Vorteil besitzt, sowohl ein weit verbreiteter, offener Standard zu sein, als auch von einer leistungsfähigen, ausgereiften Technologie zum Verarbeiten von digitalen Dokumenten unterstützt zu werden.

scheidet sich die Arbeit des Transkribenden wesentlich von Schritt zu Schritt, insbesondere was den jeweils relevanten Transkriptionsausschnitt und die erforderlichen technischen Hilfsmittel anbelangt:

Bei der Transkription der geäußerten Wörter lenkt der Transkribend in der Regel seine Aufmerksamkeit auf einen kleinen Ausschnitt der Aufnahme und konzentriert sich auf das verbale Verhalten jeweils *eines* Sprechers. Das nützlichste Hilfsmittel in diesem Schritt ist ein Werkzeug, das ein flexibles Abspielen, Anhalten, Rückspulen etc. der Aufnahme ermöglicht und es dem Transkribenden erlaubt, Transkriptionstext einzugeben und über Zeitmarken mit der Aufnahme zu verbinden. Das präzise Ordnen von Beiträgen verschiedener Sprecher hingegen erfordert eine Darstellung, die die zeitlichen Verhältnisse zwischen verschiedenen Transkriptionseinheiten in intuitiver Form wiedergibt und zu verändern erlaubt. Bei der Annotation prosodischer Eigenschaften wiederum muss dem Transkribenden der bereits vorhandene Transkriptionstext in sehr kleinen Ausschnitten und möglichst zusammen mit phonetisch basierten Visualisierungen (z.B. Spektrogramm oder graphische Darstellung der Pitch-Kontur) des zugehörigen Ausschnitts des Sprachsignals präsentiert werden. Noch mehr als beim ersten Schritt erfolgt hier eine Konzentration auf die Handlung *eines* Sprechers. Für das Festlegen von Segmentgrenzen ist dann eine Darstellungsform nützlich, die die resultierenden Segmente möglichst deutlich (etwa als eigenständige und klar abgegrenzte graphische Einheit) repräsentiert sowie unmittelbar vorausgehende und folgende Handlungen des betreffenden Sprechers leicht visuell zugänglich macht. Beim Schritt des Hinzufügens deutscher Übersetzungen schließlich tritt die Aufnahme weitestgehend in den Hintergrund, und es kommt vor allem auf eine intuitiv bearbeitbare Präsentation von Äußerung/Übersetzung-Paaren an.

Obwohl also idealerweise verschiedene Transkriptionsschritte unterschiedliche Darstellungsformen und Hilfsmittel erfordern, werden sie nach den HIAT-Konventionen (auch in ihrer überarbeiteten Fassung) alle an ein und demselben Produkt – einer Partitur – ausgeführt. Dies mag einerseits insofern vorteilhaft sein, als der Transkribend in Form der Partitur fortwährend einen intuitiv erfassbaren Überblick über die gesamte bereits vorhandene Transkription hat. Andererseits lässt sich kaum bestreiten, dass insbesondere für solche Schritte, bei denen der Transkribend sich auf kleine Ausschnitte der Aufnahme und Transkription konzentrieren muss, andere Formen der Präsentation praktischer sein können. Wie genau Transkriptionswerkzeuge verschiedene Schritte des Transkriptionsprozesses besser unterstützen können, indem sie ihre Benutzerschnittstelle an den jeweils im Vordergrund stehenden Aspekt anpassen, bliebe zu untersuchen. Es scheint mir aber sehr wahrscheinlich, dass auf diese Weise das Transkribieren wesentlich einfacher und effizienter gestaltet werden könnte und die Werkzeuge dadurch der immer wieder geforderten "Benutzerfreundlichkeit" (vgl. auch das einleitende Zitat von Linke et al.) näher kommen könnten.

Voraussetzung dafür wäre neben der im vorangehenden Abschnitt beschriebenen Trennung von Inhalt und Darstellung von Transkriptionen eine genauere Analyse und Dokumentation der Transkription als Prozess. Gesprächsanalytische Transkriptionskonventionen wären der geeignete Ort für eine solche Dokumentation.

## 5. Theorieabhängigkeit von Transkriptionskonventionen

Wie in der Literatur immer wieder beschrieben (erstmal ausführlich bei Ochs 1979) lassen sich Transkriptionskonventionen nur dann sinnvoll formulieren und verstehen, wenn eine Vorstellung darüber besteht, wer die Transkription für welchen Untersuchungszweck einsetzen soll. In bestehenden Konventionen ist diese Zweckgebundenheit deutlich erkennbar. So werden in HIAT viele Transkriptionsregeln über ihre Relevanz für eine funktional motivierte Untersuchung gesprochener Sprache gerechtfertigt (vgl. z.B. die Einheit 'Äußerung' oder das Phänomen 'Ankündigung'), und große Teile der Konventionen sind nur vor dem Hintergrund der funktional-pragmatischen Theorie überhaupt zu verstehen. Im scheinbaren Kontrast dazu versteht sich GAT als "schulenübergreifendes" (Selting 2001:1061) System, das "keine stark theoriegebundenen Vorannahmen macht" (Selting et al. 1998:92). Gleichwohl gibt es auch dort Einheiten der Transkription, die kaum ohne Rückbezug auf eine spezifische Theorie zu verstehen und verwenden sind (etwa die 'Phrasierungseinheit'), und zumindest in den Vorgaben für die Feintranskription wird deutlich, dass GAT der Beschreibung prosodischer Merkmale ein besonderes Gewicht zukommen lässt und insofern durchaus auf einen spezifischen Analysezweck ausgerichtet ist.

Grundsätzlich ist diese Theorie- und Zweckgebundenheit von Transkriptionskonventionen kaum vollständig zu vermeiden, und sie ist wohl auch der Hauptgrund dafür, dass innerhalb der Gesprächsforschung überhaupt mehrere Systeme neben- und in Konkurrenz zueinander existieren. Mit Hinblick auf die Konstruktion von Instrumenten zur computergestützten Transkription wäre es jedoch wünschenswert, theorieunabhängige (oder theorieübergreifend unkontrovers) Bereiche der Konventionen deutlicher von solchen zu trennen, die stärker auf einen bestimmten Analysezweck oder eine bestimmte Theorie ausgerichtet sind.

So unterscheiden sich etwa die Vorgaben zur orthographischen bzw. literarischen Transkription nur unwesentlich zwischen den verschiedenen Systemen. Ähnliches gilt etwa für die Transkription von Pausen oder parasprachlichen Merkmalen (z.B. Lachen, Husten etc.). Wären diese systemübergreifenden Gemeinsamkeiten in den Konventionen als solche ausgezeichnet, würde das nicht nur dem Transkribenden einen Vergleich zwischen den Systemen erleichtern. Es würde darüber hinaus auch die Konstruktion von Transkriptionssystemen ermöglichen, die einerseits (nicht theoriespezifische) Grundfunktionen zur Verfügung stellen, welche über verschiedene Systeme hinweg genutzt werden können. Diese Grundfunktionen könnten dann durch (theoriespezifische) Erweiterungsfunktionen ergänzt werden, die für die theoretischen Herangehensweisen der einzelnen Systeme optimiert sind. Die im einleitenden Zitat formulierte Kritik, dass "die meisten [Transkriptionssysteme] nicht für die Transkription mit GAT verwendet werden [können]" würde sich so beispielsweise relativieren – diejenigen Bereiche von GAT, die theorieübergreifend anwendbar sind, sollten sich bereits jetzt auch mit solchen Transkriptionswerkzeugen umsetzen lassen, die nicht speziell auf GAT ausgerichtet sind. Voraussetzung dafür wäre jedoch, dass diese Bereiche in den Konventionen eindeutig identifiziert werden.

## Zusammenfassung

Wie dieser Beitrag zu zeigen versucht hat, müssen Transkriptionskonventionen ihren Teil dazu beitragen, dass Systeme zur computergestützten gesprächsanalytischen Transkription die in sie gesetzten Erwartungen erfüllen können. Die wichtigsten Desiderata seien hier noch einmal in komprimierter Form aufgelistet:

- Transkriptionskonventionen sollten so formuliert sein, dass ihnen ein formales Datenmodell an die Seite gestellt werden kann. Dieses muss sowohl die einzelnen Transkriptionszeichen, als auch die verschiedenen dem Zeichen übergeordneten Einheiten der Transkription in eindeutiger Weise definieren. Ein solches Modell bildet dann die Grundlage für jegliche computergestützte Verarbeitung von Transkriptionsdaten.
- Bei der Formulierung der Konventionen sollten Inhalts- und Darstellungsaspekt der Konventionen konsequent voneinander getrennt werden. Das zugehörige Datenmodell sollte ausschließlich die Inhaltsseite der Transkription beschreiben.
- Transkriptionskonventionen (und das zugehörige Datenmodell) sollten nach Möglichkeit modular aufgebaut werden. Die Modularisierung sollte dabei die Prozesshaftigkeit des Transkribierens ebenso im Auge haben wie die unterschiedlichen Grade der Theorieabhängigkeit

Abschließend scheint es mir wichtig zu betonen, dass all diese Änderungen zwar durchaus grundlegender Natur sind, die Grundprinzipien gesprächsanalytischer Transkription jedoch weitestgehend unberührt lassen. Mit anderen Worten: die hier formulierten Desiderata beziehen sich ausnahmslos auf Bereiche, die in der gesprächsanalytischen Methodologie bislang einfach unbeachtet geblieben sind und stellen nicht den Wert etablierter Verfahren in Frage. Dies sollte es umso einfacher machen, die zusätzlichen Anforderungen bei zukünftigen Überarbeitungen gesprächsanalytischer Transkriptionskonventionen zu beachten.

## Literatur

- Deppermann, Arnulf (1999): Gespräche analysieren: eine Einführung. Opladen: Leske & Budrich.
- Du Bois, John W. / Chafe, Wallace L. / Meyer, Charles / Thompson, Sandra A (2000-2005): Santa Barbara corpus of spoken American English, Parts 1-4. Philadelphia: Linguistic Data Consortium. <http://www.linguistics.ucsb.edu/research/sbcorpus.html>
- DuBois, John (1991): Transcription design principles for spoken discourse research. In: Pragmatics 1, 71-106.
- Ehlich, Konrad / Rehbein, Jochen (1976): Halbinterpretative Arbeitstranskriptionen (HIAT). In: Linguistische Berichte 45, 21-41.
- Ehlich, Konrad / Rehbein, Jochen (1979b): Erweiterte halbinterpretative Arbeitstranskriptionen (HIAT2): Intonation. In: Linguistische Berichte 59, 51-75.
- Johansson, Stig (1995): The approach of the Text Encoding Initiative to the encoding of spoken discourse. In: Leech et al. (1995a), 82-98.

- Klein, Wolfgang / Schütte, Wilfried (2001): Transkriptionsrichtlinien für die Eingabe in DIDA. Mannheim: Institut für Deutsche Sprache.
- Linke, Angelika / Nussbaumer, Markus / Portmann Paul R./ Willi, Urs (2004): Studienbuch Linguistik. 5. erweiterte Auflage. Tübingen: Niemeyer.
- Ochs, Elinor (1979) Transcription as theory. In: Ochs, E. & Schieffelin, B. B. (ed.): Developmental pragmatics. New York: Academic Press, 43-72.
- Rehbein, Jochen (2001): Das Konzept der Diskursanalyse. In: Brinker, Klaus / Antos, Gerd / Heinemann, Wolfgang / Sager, Sven (Hrsg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. Berlin / New York: De Gruyter, 927-945.
- Rehbein, Jochen / Griebhaber, Wilhelm / Löning, Petra / Hartung, Marion / Bührig, Kristin (1993): Manual für das computergestützte Transkribieren mit dem Programm syncWRITER nach dem Verfahren der Halbinterpretativen Arbeitstranskriptionen (HIAT). Hamburg: Germanisches Seminar, Universität Hamburg.
- Rehbein, Jochen / Schmidt, Thomas / Meyer, Bernd / Watzke, Franziska / Herkenrath, Annette (2004): Handbuch für das computergestützte Transkribieren nach HIAT. Arbeiten zur Mehrsprachigkeit, Serie B (56). Hamburg: SFB Mehrsprachigkeit.
- Schmidt, Thomas (2002): Gesprächstranskription auf dem Computer – das System EXMARaLDA. In: Gesprächsforschung 3, 1-23.
- Schmidt, Thomas (2005) : Computergestützte Transkription. Modellierung und Visualisierung gesprochener Sprache mit texttechnologischen Mitteln. Frankfurt a.M.: Peter Lang.
- Schmidt, Thomas / Wörner, Kai (2005): Erstellen und Analysieren von Gesprächskorpora mit EXMARaLDA. In: Gesprächsforschung 6, 171-195.
- Schütte, Wilfried (2004): Transkriptionsrichtlinien für die Eingabe in EXMARaLDA nach DIDA-Konventionen. Mannheim: Institut für Deutsche Sprache.
- Selting, Margaret (2001): Probleme der Transkription verbalen und paraverbalen/prosodischen Verhaltens. In: Brinker, Klaus / Antos, Gerd / Heinemann, Wolfgang / Sager, Sven (Hrsg.): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. Berlin / New York: De Gruyter, 1059-1068.
- Selting, Margret / Auer, Peter / Barden, Birgit / Bergmann, Jörg / Couper-Kuhlen, Elizabeth / Günthner, Susanne / Meier, Christoph / Quasthoff, Uta / Schlobinski, Peter / Uhmann, Susanne (1998): Gesprächsanalytisches Transkriptionssystem (GAT). In: Linguistische Berichte (173), 91-122.
- Witt, Andreas (2002): Multiple Informationsstrukturierung mit Auszeichnungssprachen. XML-basierte Methoden und deren Nutzen für die Sprachtechnologie. Dissertation, Universität Bielefeld.

Dr. Thomas Schmidt  
SFB 538 'Mehrsprachigkeit'  
Max Brauer-Allee 60, 22765 Hamburg  
thomas.schmidt@uni-hamburg.de

Veröffentlicht am 11.2.2008

© Copyright by GESPRÄCHSFORSCHUNG. Alle Rechte vorbehalten.