

GAT: Aspekte der computertechnischen Umsetzbarkeit

31. Januar 2008

Thomas Schmidt (Universität Hamburg)
Arnulf Deppermann (IDS Mannheim)
Martin Hartung (IDS Mannheim)
Wilfried Schütte (IDS Mannheim)

INHALTSVERZEICHNIS

EINLEITUNG	4
WAS BEDEUTET „EIGNUNG FÜR COMPUTERGESTÜTZTE VERFAHREN“?	4
ANSATZ.....	4
MODULARISIERUNG	5
GLIEDERUNG.....	5
1. VERBALTRANSKRIPTION	6
1.1. MODELL	6
1.1.1. Grundbausteine der Transkription	6
Wörter	6
Besondere Wörter	7
Andere phonologisch beschreibbare Einheiten.....	7
Nicht phonologisch beschreibbare Einheiten.....	8
Ein- und Ausatmen	8
Pausen	9
Unverständliches	9
1.1.2. Sprecherzuordnung.....	9
1.1.3. Zeitliche Ordnung.....	10
1.1.4. Turns.....	11
1.1.5. Nichtselbstständige Einheiten.....	12
Vermuteter Wortlaut und Alternativlautung	12
1.2. VISUALISIERUNG	12
2. MINIMALTRANSKRIPTION	13
2.1. MODELL	13
2.1.1. Phrasierungseinheiten	13
2.1.2. Silben	14
2.1.3. Zeitliche Ordnung.....	14
2.1.4. Nichtselbstständige Einheiten.....	15
Haupt-/ Primärakzent.....	15
2.2. VISUALISIERUNG	15
3. BASISTRANSKRIPTION	17
3.1. MODELL	17
3.1.1. Phoneme	17
3.1.2. Nichtselbstständige Einheiten.....	17
Tonhöhenbewegung am Einheitenende	17
Schneller Anschluss („Latching“), Verschleifung	17
Dehnung	17
Interpretierende Kommentare	18
3.2. VISUALISIERUNG	18
4. FEINTRANSKRIPTION	19
4.1. MODELL	19
4.1.1. Nichtselbstständige Einheiten gemäß GAT 1.....	19
Neben-/Sekundärakzent und extra starker Akzent.....	19
Auffällige Tonhöhen sprünge	19
Verändertes Tonhöhenregister.....	19
Akzenttonhöhenbewegungen.....	19
Lautstärke- und Sprechgeschwindigkeitsveränderungen.....	19
4.1.2. Nichtselbstständige Einheiten gemäß GAT 2.....	19
4.2. VISUALISIERUNG	19
ZUSAMMENFASSUNG	21
LITERATUR	22
ANHANG: SCHEMATISCHE ÜBERSICHTEN	23
1. TRANSKRIPTIONSSTUFEN (NACH THEORIEABHÄNGIGKEIT, KOMPLEXITÄT DES DATENMODELLS)	23
2. EINHEITEN	23
2.1. SELBSTSTÄNDIGE EINHEITEN.....	23

2.2. NICHT-SELBSTSTÄNDIGE EINHEITEN	23
2.2.1. Verbaltranskription	23
2.2.2. Minimaltranskription.....	23
2.2.3. Basistranskription.....	23
2.2.4. Feintranskription:.....	23
3. ZEICHEN.....	24
3.1. Zeichen, die selbstständige Einheiten beschreiben.....	24
3.2. Zeichenketten, die selbstständige Einheiten beschreiben	24
3.3. Grenzmarkierer	24
3.4. Zeichen und Zeichenketten, die zur Visualisierung von nicht-selbstständigen Einheiten verwendet werden	25
Verbaltranskription.....	25
Minimaltranskription	25
Basistranskription	25
Feintranskription	25

Einleitung

Dieser Text analysiert die GAT-Konventionen (Selting et al. 1998) hinsichtlich ihrer Eignung für computergestützte Verfahren der Erstellung, Auswertung und Archivierung von Transkriptionskorpora. Ziel ist, diesbezügliche Unzulänglichkeiten zu identifizieren und Vorschläge zu machen, wie diese in einer überarbeiteten GAT-Version behoben werden können. Da GAT derzeit auch in anderen Teilbereichen (Prosodie, Multimodalität, literarische Umschrift) überarbeitet bzw. ergänzt wird, hat diese Analyse zwangsläufig nur einen vorläufigen Status – sie sollte möglichst vor der Veröffentlichung der überarbeiteten Konventionen an den dann vorliegenden Entwürfen wiederholt werden. Es ist jedoch davon auszugehen, dass sich die Grundprobleme und deren mögliche Lösungen mit der Überarbeitung nicht substantiell ändern.

Was bedeutet „Eignung für computergestützte Verfahren“?

Eine überarbeitete Version der GAT-Konventionen soll der Tatsache Rechnung tragen, dass gesprächsanalytische Transkription heute fast ausnahmslos mit Hilfe des Computers vorgenommen wird. Von computergestützten Methoden wird erwartet, dass sie erstens die Arbeit von Transkribenden und Gesprächsforschern erleichtern, zweitens aber auch neue Möglichkeiten der Darstellung und Analyse eröffnen. Transkriptionskonventionen müssen ihren Teil dazu beitragen, damit dies möglich wird. Die bestehende Version der GAT-Konventionen ist in dieser Hinsicht unzulänglich. „Eignung für computergestützte Verfahren“ würde bedeuten:

- Ein Transkribend kann anhand der Konventionen fortgeschrittene Transkriptionssoftware nutzen, um damit GAT-konforme Transkriptionen anzufertigen.
- Ein Softwareentwickler kann anhand der Konventionen (und evtl. begleitender technischer Dokumentation) Transkriptionssoftware schreiben, die das Transkribieren nach GAT effizient und zuverlässig unterstützt.
- Gesprächsforscher können auf Grundlage der Konventionen verlässliche computergestützte Suchen auf Transkriptionskorpora ausführen. „Verlässlich“ bedeutet dabei insbesondere, dass Unterselektion (d.h. das versehentliche Nicht-Auffinden gesuchter Phänomene) und Überselektion (d.h. das unerwünschte Auffinden nicht gesuchter Phänomene) so weit wie irgend möglich vermieden werden können (vgl. dazu z.B. Edwards 1995).
- Die Konventionen schaffen die Grundlage für die Nachhaltigkeit, d.h. für die langfristige Verwendbarkeit und Austauschbarkeit (zwischen Forschern, zwischen Betriebssystemen, zwischen Softwarewerkzeugen) von Transkriptionskorpora, die in digitalen Archiven aufbewahrt und zugänglich gemacht werden.

Ansatz

Die Analyse der GAT-Konventionen wird nach dem in Schmidt (2005) verwendeten Ansatz vorgenommen, der hier nicht im Detail erläutert werden soll. Die wichtigsten Prinzipien sind im Folgenden stichwortartig aufgeführt:

- Trennung von Modell- und Visualisierungsaspekt
- Dreiebenenarchitektur (logische, physikalische und Anwendungsebene) der Datenverarbeitung
- Analyse des Modellaspekts nach dem Entity-Relationship-Paradigma
- Unterscheidung zwischen selbstständigen und nicht-selbstständigen Einheiten der Transkription
- Unterscheidung zwischen verschiedenen Typen symbolischer Beschreibung: Transkription, Deskription, numerische Beschreibung, geschlossene Liste

Modularisierung

Die derzeitige Fassung der GAT-Konventionen unterscheidet zwischen Basis- und Feintranskript. Auf dem GAT-Treffen im Oktober 2007 wurde vorgeschlagen, diese Unterteilung zu verfeinern und ein oder zwei weitere Transkriptstufen unterhalb des Basistranskripts einzuführen. Der vorliegende Text folgt diesem Vorschlag, indem er vier Transkriptionsstufen vorschlägt:

1. die Verbaltranskription
2. die Minimaltranskription (oder Roh- oder Arbeitstranskription)
3. die Basistranskription
4. die Feintranskription

Diese Unterteilung orientiert sich am Transkriptionsprozess, indem sie die verschiedenen Einheiten der Transkription so ordnet, dass die Transkriptionsstufen nacheinander und ohne Revision vorheriger Stufen angefertigt werden können (Stichwort: „Zwiebelprinzip“). Dadurch können nicht nur die Konventionen bzw. deren Analyse übersichtlicher gehalten werden, sondern es wird auch dem Umstand Rechnung getragen, dass Transkriptionen mit zunehmendem Feinheitegrad theorieabhängiger werden und komplexere Datenmodelle erfordern. Die Modularisierung trägt somit erstens dazu bei, dass sich schulenübergreifend unstrittige Bereiche der Konventionen als solche identifizieren lassen und Transkriptionen damit für einen größeren Anwenderkreis nutzbar werden. Zweitens erleichtert sie durch ihre Prozessorientiertheit und die Abstufung bzgl. der Komplexität des Datenmodells die Konstruktion von Software-Werkzeugen zur Transkription.

Gliederung

Die vier Abschnitte dieses Textes beschreiben nacheinander die Stufen der Verbaltranskription, der Minimaltranskription, der Basistranskription und der Feintranskription. Jeder Abschnitt ist weiter unterteilt in einen Modell- und einen Visualisierungsteil. Im Modellteil werden die verschiedenen selbstständigen und nicht-selbstständigen Einheiten nacheinander beschrieben und kommentiert (die Besprechungen von Detailproblemen sind dabei vom übrigen Text abgesetzt). Außerdem werden hier Fragen der Sprecherzuordnung und der zeitlichen Ordnung behandelt. Im Visualisierungsteil werden einerseits Fragen behandelt, die die Visualisierung für die Ausgabe betreffen. Andererseits geht es hier auch um Probleme, die bei der Eingabe einer GAT-Transkription in einem Text-Editor auftreten können. Fragen, die sich nur auf die Publikation von Transkriptausschnitten beziehen (z.B. die Zeilennummerierung, das Einrücken und anderweitige Kenntlichmachen von relevanten Stellen des Transkripts) werden in diesem Text zunächst nicht behandelt.

1. Verbaltranskription

Auf der Stufe der Verbaltranskription werden nur solche Einheiten berücksichtigt, die weitestgehend unabhängig von spezifischen linguistischen Theorien und Fragestellungen sind. Im Idealfall sollte dieser Teil der Konventionen also auch in anderen Zusammenhängen als der Gesprächsforschung verwendbar sein. Die Datenstrukturen sind auf dieser Stufe so einfach, dass sie sich sowohl mit allen gängigen Transkriptionswerkzeugen (insb. Praat, ELAN, EXMARALDA, CLAN), als auch mit einem Texteditor erzeugen und verarbeiten lassen.

1.1. Modell

1.1.1. Grundbausteine der Transkription

Wörter

Die GAT-Konventionen erläutern die Einheit „Wort“ nicht. Wortschreibungen werden somit nach den Vorgaben der Orthographie bzw. der literarischen Umschrift¹ vorgenommen, allerdings mit genereller Kleinschreibung. Wortgrenzen werden ebenfalls nach den Vorgaben der Orthographie gesetzt. Bei manchen Assimilationen fordern die Vorgaben zur literarischen Umschrift jedoch, dass zwei orthographische Wörter in einer transkribierten Form zusammengefasst werden (z.B. „vonne“ für ‚von der‘ oder „schauma“ für ‚schauen wir‘, S.8). Die Vorgaben sind weitestgehend unproblematisch. Zu überlegen wäre aber:

- Sollen die orthographischen Wortgrenzen nicht auch bei Assimilationen beibehalten werden (also: „schau ma“ statt „schauma“)? Dies würde eine Lemmatisierung und die Durchsuchbarkeit der Transkriptionen vereinfachen, ist aber u.U. schwer in eine allgemein anwendbare Regel zu packen (wo liegt z.B. die orthographische Wortgrenze in „hasse“ – ‚hast Du‘?).
- Bindestrichkomposita (‚Bild-Zeitung‘) sind problematisch, weil der Bindestrich auch zur Visualisierung einer gleich bleibenden Tonhöhenbewegung verwendet wird. Dies könnte umgangen werden, indem grundsätzlich eine Schreibung ohne Bindestrich (also ‚bildzeitung‘) gefordert wird.
- Vergleichbare Probleme kann es auch mit dem Apostroph geben, etwa wenn ein Sprecher englische Einschübe in seiner Rede hat (z.B. ‚Mein Lieblingslied ist I’m a believer‘ oder ‚wir sind zum Speaker’s Corner gegangen‘), siehe dazu auch unten unter „Glottalverschluss“.
- Eingedeutschte Ausdrücke oder fremdsprachliche Eigennamen, die von Erweiterungen des lateinischen Alphabets Gebrauch machen, die sonst nicht im Deutschen gebraucht werden (z.B. ‚peu à peu‘, ‚Charité‘, ‚Citroën‘, etc.) sind gleichfalls problematisch, weil sie die Menge von Zeichen, aus denen Wörter aufgebaut sind, in potentiell unvorhersagbarer Weise erweitern. Auch hier könnte grundsätzlich gefordert werden, diese Formen ohne Verwendung der problematischen Zeichen (also ‚peu a peu‘, ‚charite‘, ‚citroen‘) zu transkribieren. Zumindest sollte aber eine Liste solcher Zeichen erstellt werden, anhand derer potentielle Konflikte mit anderen Transkriptionszeichen erkannt werden können.

¹ Ob, an welchen Stellen und nach welchen Regeln statt der Standardorthographie die literarische Umschrift eingesetzt wird, hat weit reichende Auswirkungen auf die computergestützte Verarbeitbarkeit von Transkriptionen, insbesondere auf ihre zuverlässige Durchsuchbarkeit. Gleiches gilt für die Entscheidung über die generelle Kleinschreibung. Es gibt verschiedene Möglichkeiten, diesbezüglichen Problemen zu begegnen (z.B. eine Lemmatisierung). Diese Möglichkeiten können besser diskutiert werden, wenn die Vorgaben zur literarischen Umschrift überarbeitet sind. Die Thematik wird daher in diesem Text nicht weiter behandelt.

Besondere Wörter

Besondere Wörter, sind solche, deren Orthographie nicht dem sonst gültigen phonographischen Prinzip folgt. Dazu zählen mindestens:

- Zahlen
- Abkürzungen
- Buchstabierungen

Die GAT-Konventionen machen zu diesen besonderen Wörtern bislang keine Vorgaben. Diese wären aber dringend nötig, weil solche Ausdrücke sonst nicht zuverlässig verarbeitet werden können.

- Zahlen sollten in jedem Falle aus- (also nicht als Ziffernfolgen) geschrieben werden, nicht zuletzt, weil nur so Unterschiede wie ‚zweiundzwanzig‘ vs. ‚zwoundzwanzig‘ oder ‚einhundert-eins‘ vs. ‚hunderteins‘ erfasst werden können.
- Abkürzungen enthalten nach den Vorgaben der Orthographie oft einen Punkt (z.B. ‚St. Pauli‘) der aber bereits anderweitig (als Markierung einer tief fallenden Tonhöhenbewegung) eingesetzt wird. Auch hier ist es am konsequentesten, einfach ein Ausschreiben der Ausdrücke zu fordern (also ‚sankt pauli‘).
- Bei Buchstabierungen (‚Müller schreibt sich m-ü-l-l-e-r‘) und buchstabierten Abkürzungen (‚DDR‘, ‚Dr. h.c.‘, ‚GmbH‘) sind zwei Regelungen denkbar: Entweder werden nur die einzelnen Buchstaben (in Kleinschreibung, getrennt durch Leerzeichen) verwendet – also ‚Müller schreibt sich m ü l l e r‘ oder ‚hier sieht es aus wie in der d d r‘ –, oder die Buchstabierungen werden nach dem Prinzip der literarischen Umschrift vorgenommen – also ‚Müller schreibt sich em üh el el eh er‘ bzw. ‚hier sieht es aus wie in der de de er‘. Letztere Lösung wäre konsistenter mit dem Rest der Konventionen. Beide Lösungen führen neue Verwechslungsmöglichkeiten ein. U.U. könnte man Buchstabierungen auch gesondert kenntlich machen, z.B. durch einen vorangestellten Unterstrich, also ‚_d _d _r‘ oder ‚_de _de _er‘).

Andere phonologisch beschreibbare Einheiten

Neben Wörtern werden auch einige andere Einheiten nach dem phonographischen Prinzip beschrieben, nämlich:

- Verzögerungssignale
- Glottalverschluss bei Abbruch
- Lachen
- Rezeptionssignale

Diese sind auf Seite 11f durch Beispiele geeigneter „orthographischer Annäherungen“ beschrieben.

- Allgemein wären zusätzlich zu den Beispielen ein paar Erläuterungen wünschenswert.
- Das Apostroph-Zeichen, das für die Darstellung des Glottalverschlusses verwendet wird, kann potentiell Probleme verursachen, wenn bei englischen Passagen ein Saxon Genitive oder eine Form wie ‚hasn’t‘ gemäß der Orthographie transkribiert würde. Da englische Passagen u.U. häufiger vorkommen, sollte dieser Fall explizit geregelt werden, z.B. indem hier eine Schreibweise ohne Apostroph gefordert wird.
- Das Beispiel „so(h)o“ für „Lachpartikel beim Reden“ steht im Konflikt zur Vorgabe für das Transkribieren schwer verständlicher Wortteile (Beispiel „al(s)o“, Seite 16). Hier muss eine alternative Regelung gefunden werden.
- Die Regelung, zweisilbige Rezeptionssignale mit einem Gleichheitszeichen an der Silbengrenze zu transkribieren („hm=hm“, „nei=ein“) führt dazu, dass diese Einheiten nicht von Verschleifungen zweier Wörter („und=äh“) zu unterscheiden sind. M.a.W.: bei einer automa-

tischen Segmentierung auf Grundlage einer Textdatei würden zweisilbige Rezeptionssignale als zwei Wörter segmentiert. Falls das nicht gewünscht ist, müsste auch hier eine alternative Regelung gefunden werden.

Nicht phonologisch beschreibbare Einheiten

Nicht phonologisch beschreibbare Einheiten werden in den Konventionen pauschal unter dem Stichwort „nonverbale Handlungen und Ereignisse“ behandelt (S. 15). Es werden außer einigen Beispielen („schnieft“, „hustet“) keine Vorgaben zu deren Beschreibung gemacht.

- Auf der Ebene der Verbaltranskription sollten grundsätzlich nur solche non-verbale Handlungen und Ereignisse berücksichtigt werden, die nicht parallel zu einer verbalen Handlung des gleichen Sprechers erfolgen. Die Transkription parallel verlaufender Handlungen eines Sprechers erfordert Datenstrukturen und Transkriptionsanweisungen einer höheren Komplexität, die – wenn irgend möglich – erst auf einer höheren Transkriptionsstufe (der Basistranskription) eingeführt werden sollten (aus dem gleichen Grund sollten auch interpretierende Kommentare erst in der Basistranskription Anwendung finden).
- Fälle, in denen eine non-verbale Handlung eine verbale teilweise überlappt, müssen geregelt werden. Gemäß dem vorherigen Punkt sollte hier nur der nicht parallel zu Verbalem stattfindende Teil transkribiert werden, also etwa: „sag ich doch ((hustet))“, auch wenn das Husten bereits beim Wort „ich“ einsetzt.
- Die Konventionen machen keinerlei Vorgaben, nach denen der Transkribend entscheiden könnte, ob er eine gegebene Handlung oder ein Ereignis in der Transkription berücksichtigen soll oder nicht. Erfahrungsgemäß kann das dazu führen, dass auch solche Phänomene beschrieben werden, die für die Analyse irrelevant sind („Kaffeemaschine blubbert“). Dies tritt verstärkt auf, wenn von einer Video-Aufnahme transkribiert wird. Bei der Verarbeitung von Transkriptionen kann eine solche übertriebene Berücksichtigung von Nonverbalem hinderlich sein. Die Konventionen sollten einen Begriff wie „kommunikative Relevanz“ einführen, anhand dessen Entscheidungen über die (Nicht-)Berücksichtigung eines non-verbale Ereignisses bei der Transkription getroffen werden können.
- Darüber hinaus wäre es wünschenswert, die Konventionen in diesem Bereich etwas weniger stichwortartig zu gestalten. Es könnten zumindest einige genauere Vorgaben zu den zu verwendenden Beschreibungen (etwa: „möglichst als Verb in der dritten Person Singular“) gemacht und eine grobe Unterteilung der non-verbale Handlungen und Ereignisse (etwa: „vokal vs. nicht vokal“, „akustisch vs. visuell wahrnehmbar“, siehe dazu auch die TEI-Richtlinien) vorgenommen werden und für jeden Teilbereich ein paar zusätzliche Erläuterungen und Beispiele gegeben werden.
- Die Beschreibungen, die für Nonverbales verwendet werden, lassen – im Gegensatz zur Transkription von Verbalem – für sich genommen keinerlei Aufschluss über die Dauer der Handlung oder des Ereignisses zu. Dies kann erfahrungsgemäß beim Lesen von Transkriptionen zu Missverständnissen führen (etwa: bezeichnet „weint“ eine Handlung von einer halben Sekunde oder von zwei Minuten?). U.U. wäre es sinnvoll, eine Vorgabe zu machen, wie die Dauer einer Handlung zu beschreiben ist, z.B. „((weint, 10s))“.

Ein- und Ausatmen

Ein- und Ausatmen wird nach den derzeitigen Konventionen erst in der Feintranskription berücksichtigt (S. 22). Die Vorgabe lautet „je nach Länge“, die Zeichenketten .h, .hh, .hhh bzw. h, hh, und hhh zu verwenden.

- Das Ein- und Ausatmen ist die einzige Einheit in der Feintranskription, die selbstständig ist (sich also nicht auf anderes, bereits Transkribiertes bezieht). Nach dem Prinzip, dass alle Grundbausteine der Transkription Teil der Verbaltranskription sein sollten, gehört auch das Ein- und Ausatmen auf diese Transkriptionsstufe. Es stellt sich in diesem Zusammenhang auch die Frage, was der Transkribend bei der Ersttranskription an Stelle des Ein- und Ausatmens setzen würde (eine Pause?), wenn er selbiges erst bei der Feintranskription berücksichtigen soll.

- Der Punkt, der zum Darstellen des Einatmens verwendet werden soll, steht im Konflikt mit dem Punkt, der eine tief fallende Tonhöhenbewegung beschreibt. Ein alternatives Zeichen wäre hier empfehlenswert.
- Die Angaben „je nach Dauer“ bzw. „sehr kurz“, „normal lang“ und „deutlich länger“ sind im derzeitigen Konventionstext nicht sinnvoll anwendbar. Es fehlt die Angabe einer absoluten Dauer oder einer relativen Bezugsgröße, nach denen sich diese Formulierungen interpretieren lassen.
- Es sollte außerdem deutlich gemacht werden, welche Art von Ein- und Ausatmen überhaupt zu berücksichtigen ist. Wahrscheinlich geht es um hörbares Ein- und Ausatmen, das nicht parallel zu Gesprochenem verläuft.

Pausen

Nach der derzeitigen Fassung der Konventionen können Pausen geschätzt oder gemessen werden. Für geschätzte Pausen ist neben der Angabe der Dauer als Zahl alternativ auch eine Verwendung der Symbolketten (.), (-), (--), (---) vorgesehen (S. 10). Die Vorgaben an sich sind grundsätzlich unproblematisch. Ihre Darstellung ist aber relativ verwirrend:

- Wie ist die Formulierung „kurze, mittlere und längere geschätzte Pause von 0.25 – 0.75 Sek. Dauer; bis zur Pausendauer von ca. 1 Sek.“ zu verstehen? Ist 0.75 Sek. oder 1 Sek. die Obergrenze?
- Sind Mikropausen damit Pausen, deren geschätzte Länge weniger als eine Viertelsekunde beträgt?
- Die Ausführungen in der Fußnote zum Unterschied zwischen absolut gemessener Pausendauer und Messung der Pausendauer in Relation zum Rhythmus sind kaum mit dem Konventionstext in Einklang zu bringen. Wie würde ein Transkribend eine Pause transkribieren, die er „in Relation zur Dauer umliegender akzentuierter Silben [gemessen]“ hat? Müssten dafür nicht statt oder zusätzlich zu den absoluten Orientierungswerten im Konventionstext auch Maße für solche Relationen spezifiziert werden (etwa: ‚Mikropause: ca. ein Drittel so lang wie eine vorangehende akzentuierte Silbe‘ etc.)?

Unverständliches

Nach den derzeitigen Vorgaben werden unverständliche Passagen als „()“ transkribiert und zwar „je nach Länge mit unterschiedlich viel Leerraum“. Beim GAT-Treffen im Oktober 2007 wurde bereits vorgeschlagen, dies zu präzisieren und eine Schreibweise zu verwenden, die die Anzahl der unverständlichen Silben wiedergibt, z.B. „xxx xxx xxx“, wobei jedes „xxx“ eine Silbe repräsentiert. Für die computergestützte Verarbeitung wäre eine solche Darstellung besser. Allerdings stellt sich die Frage, wie zu verfahren ist, wenn der Transkribend auch die Anzahl der Silben nicht feststellen kann. Sinnvoll in diesem Falle wäre eine Angabe wie „((unverständlich, 2.5s))“, womit solche Passagen analog zu nonverbalen Ereignissen behandelt würden.

1.1.2. Sprecherzuordnung

Vorgaben für die Sprecherzuordnung werden explizit nur bei den Pausen gemacht (S. 10). Außerdem wird die Sprecherzuordnung auch unter „Allgemeine Struktur des Transkripts“ (S. 5) behandelt.

- Es wird keine Vorgabe gemacht, wie mit unbekanntem Sprechern zu verfahren ist. Andere Konventionen machen hier Vorgaben wie „nicht identifizierbare Sprecher werden mit X1, X2, ... bezeichnet, bzw. mit F1, F2, ... und M1, M2, ..., sofern sie eindeutig weiblich oder männlich sind“.
- Bei den Vorgaben zur Transkription von Nonverbalem gibt es keine Erläuterungen zur Sprecherzuordnung. Es sollte hier auf jeden Fall erwähnt werden, dass manche solcher Handlungen keinem Sprecher zugeordnet werden können (z.B. „Telefon klingelt“, „Aufnahmegerät piepst“), vielleicht auch, wie zu verfahren ist, wenn ein Ereignis mehreren Sprechern zugeord-

net werden könnte („Erwin und Klaus schütteln sich die Hand“ vs. „schüttelt Klaus die Hand“ als Handlung, die dem Sprecher Erwin zugeordnet ist).

- Bei den Vorgaben zur Sprecherzuordnung von Pausen bleibt unklar, ob es grundsätzlich möglich ist, Pausen zwischen Sprecherwechseln (also am Turnbeginn oder -ende) einem der betreffenden Sprecher zuzuordnen, oder ob solche Pausen grundsätzlich keinem Sprecher zugeordnet werden.

1.1.3. Zeitliche Ordnung

Eine klare zeitliche Ordnung der Transkriptionseinheiten ist entscheidend für sehr viele Aufgaben, die mit Hilfe des Computers vorgenommen werden. Gleichzeitig ist dies der Bereich, in dem die jetzige Fassung der GAT-Konventionen die meisten Unzulänglichkeiten aufweist, und zwar in zweierlei Hinsicht:

Erstens machen die Konventionen überhaupt keine Angaben darüber, wie per absoluter Zeitangabe eine Verknüpfung zwischen Transkription und Aufnahme herzustellen ist. Gerade diese Verknüpfung ist aber eine entscheidende Hilfe für das effiziente Transkribieren und für die Analyse.

Zweitens gehen die Konventionen grundsätzlich von einer sequentiellen Struktur der Transkription aus und behandeln Gleichzeitigkeit als Ausnahme von der Regel. Diese Ausnahmen werden unter dem Stichwort „Überlappungen und Simultansprechen“ (S. 8, teilweise auch noch unter „Nonverbale Handlungen und Ereignisse“ sowie „Interpretierende Kommentare“, S. 15f) behandelt. Es werden jeweils Beispiele gegeben, wie Überlappungen mit Hilfe von geeigneten Klammerungen im Transkript zu repräsentieren sind. Diese Beispiele sind jedoch nicht uneingeschränkt verallgemeinerbar. Sie behandeln jeweils einfache Fälle, in denen eine i.d.R. kurze verbale Handlung eines Sprechers eine längere verbale Handlung eines anderen Sprechers überlappt. Wie zu verfahren ist, wenn sich die Beiträge mehrere Sprecher überlappen, wenn sich Nonverbales mit Verbalem des gleichen oder eines anderen Sprechers überlappt, oder gar wenn diese beiden Phänomene auf einmal auftreten, wird nicht so spezifiziert, dass es für eine computertechnische Umsetzung verwendet werden könnte. Weitere Unklarheiten betreffen die Rolle von Interpunktionszeichen in überlappenden Passagen (etwa: „wird ein Fragezeichen, das eine Tonhöhenbewegung repräsentiert, innerhalb oder außerhalb der Klammern gesetzt?“) und den Präzisionsgrad der Markierung von Überlappungen (etwa: „Werden auch dann nur Beginn und Ende der Überlappung markiert, wenn parallele Passagen zweier Sprecher jeweils mehrere Phrasierungseinheiten umfassen?“). Außerdem führen die Regelung, die Endpunkte von Überlappungen ggf. unspezifiziert zu lassen, und die Möglichkeit, überlappende Beiträge zusätzlich durch Einrückungen zu verdeutlichen, zu Datenstrukturen, die weder für den Leser noch für einen Computer eindeutig zu interpretieren sind.

Es wird daher vorgeschlagen, hinsichtlich der zeitlichen Ordnung von Transkriptionseinheiten grundsätzlich anders zu verfahren. Folgendes soll dabei gelten:

- Von jeder Stelle der Transkription kann über die Angabe eines absoluten Zeitwertes in die Aufnahme verwiesen werden. Damit diese Angabe über verschiedene Medienformate hinweg nutzbar bleibt, sollte sie immer in Zeiteinheiten (Sekunden), nicht in Frames, Kassettenzählwerkspositionen o.Ä. vorgenommen werden.
- Alle zeitgemäßen Transkriptionswerkzeuge haben Mechanismen, um solche absoluten Zeitwerte in der Transkription zu speichern und sie nutzbar zu machen. Für den Fall, das mit einem einfachen Texteditor transkribiert werden soll, müsste aber geregelt werden, in welcher Form diese absoluten Zeitwerte in den Text integriert werden können. Ein Vorschlag wäre, die Angabe vor der Einheit, die an diesem Zeitpunkt beginnt, vorzunehmen und durch geschweifte Klammern kenntlich zu machen, z.B. „sag {23.3}ich doch“. Eventuell könnte eine zusätzlichen Position vor jedem Turnbeginn vorgesehen werden, z.B. „{58.5} KLA: guten tag“. Es wäre dann auch festzulegen, in welchem Format die Zeitangabe zu notieren ist (z.B. „{1:02.3}“ vs. „{62.03}“).

- Auf der Ebene der Verbaltranskription wird so transkribiert, dass sich die o.g. Grundbausteine der Transkription prinzipiell nicht überlappen, wenn sie dem gleichen Sprecher zugeordnet sind.
- Transkribenden werden angehalten, in der Verbaltranskription mindestens bei jedem Sprecherwechsel bzw. jeweils am Turnbeginn und -ende eine absolute Zeitmarke einzufügen. Damit sind Überlappungen zunächst implizit über die absolute Zeitzuordnung von Turns repräsentiert.
- Die explizite Transkription von Überlappungen, also das Markieren von Start- und Endpunkt der Überlappung, erfolgt erst auf der Ebene der Minimaltranskription (näheres siehe dort).
- Der Transkriptionstext sollte den Leser darauf aufmerksam machen, dass die Methode der Klammerdarstellung bei freien Gesprächen mit mehreren Teilnehmern und bei verstärkter Berücksichtigung von Nonverbalem an ihre Grenzen stößt. Für diese Fälle sollte auf die Möglichkeit einer Partitur-Darstellung und auf Transkriptionsinstrumente für Mehrebenen-Annotationen (also ELAN, Praat, EXMARaLDA etc.) verwiesen werden.

Außerdem bestehen für die derzeitige Fassung der Konventionen folgende Detailprobleme:

- Es bleibt unklar, wie zu verfahren ist, wenn sich eine Gesprächspause mit einem non-verbalem Ereignis überlappt. Sollen in diesem Falle sowohl Pause als auch Ereignis oder nur eines von beidem beschrieben werden?
- Ebenso bleibt unklar, wie zu verfahren ist, wenn zwei non-verbale Ereignisse gleichzeitig stattfinden. Nach dem Prinzip, dass in der Verbaltranskription ein und demselben Sprecher zugeordneten Einheiten einander nicht überlappen sollten, müssten diese als eine Einheit (also „nickt und steht auf“, „steht nickend auf“ o.Ä. statt „nickt“ und „steht auf“) beschrieben werden.

1.1.4. Turns

Der Begriff „Turn“ wird an verschiedenen Stellen der Konventionen verwendet, allerdings ohne irgendwo eindeutig definiert zu sein. Dabei ist jedoch klar, dass er in den Konventionen eher in einem „naiven“ als in einem gesprächsanalytisch elaborierten Sinne gebraucht wird, also etwa in dem Sinne „ein Turn ist alles, was ein Sprecher tut zwischen zwei Abschnitten, in denen er nichts tut“ und ohne die Bezugnahme auf komplexere gesprächsanalytische Kategorien. Aufbauend auf den vorherigen drei Abschnitten ist dies gleichbedeutend mit „ein Turn ist die Gesamtheit aller Transkriptions-Grundbausteine, die ein- und demselben Sprecher zugeordnet sind und die zeitlich unmittelbar aufeinander folgen“. Dies ist eine Definition, mit der sich hinsichtlich computergestützter Verfahren sehr gut arbeiten lässt, denn sie erlaubt es, die Einheit Turn aus rein formalen Eigenschaften anderer Einheiten automatisch abzuleiten. Wenn diese Definition allgemein als akzeptabel angesehen wird, stellen sich folgende Detailfragen:

- Der Transkriptionstext sollte diese Definition enthalten und möglichst deutlich machen, dass dies ein im oben beschriebenen Sinne „naives“ Verständnis des Begriffs „Turn“ ist.
- Auf der obersten Hierarchieebene bestünde eine Transkription damit aus einer Menge von Turns, die jeweils einem Sprecher zugeordnet sind, und einer Menge von Einheiten (Pausen und Nonverbales), die keinem Sprecher zugeordnet sind. Letztere müssten ebenfalls benannt werden.
- Gleichfalls wäre zu klären, ob isolierte nonverbale Handlungen, die einem Sprecher zugeordnet sind, ebenfalls als „Turns“ zu bezeichnen sind, oder ob dafür eine gesonderte Bezeichnung eingeführt werden soll.
- Dem Transkribenden muss klargemacht werden, dass die Entscheidung, ob eine Pause einem Sprecher zugeordnet wird oder nicht, oft gleichbedeutend ist mit der Entscheidung, ob an dieser Stelle ein Turn endet oder nicht

1.1.5. Nichtselbstständige Einheiten

Vermuteter Wortlaut und Alternativlautung

Auf der Ebene der Verbaltranskription sind lediglich zwei nicht-selbstständige Einheiten relevant:

- die Annotation „vermuteter Wortlaut“
- die Annotation „Alternativlautung“

Dabei beziehen sich beide Annotationen auf i.d.R. die Einheit „Wort“, es wird jedoch ein Beispiel gegeben, in dem sich die Vermutung lediglich auf einen Laut innerhalb eines Wortes bezieht, nämlich „al(s)o“.

- Die bestehende Vorgabe, schwer verständliche Wörter in runde Klammern zu setzen und innerhalb dieser Klammern optional Alternativlautungen hinter einem Schrägstrich zu nennen, z.B. „(welche/solche)“, ist bzgl. der computergestützten Verarbeitung unproblematisch. Es wäre aber wünschenswert, zumindest auf der Ebene der Verbaltranskription keine Vermutungen unterhalb der Wortebene zuzulassen. Für das genannte Beispiel könnte dann alternativ „(also/alo)“ transkribiert werden.
- Prinzipiell können sich Vermutungen und Alternativlautungen auch auf mehrere Wörter beziehen, z.B. „(die Glorreichen Sieben/das Rohr neigt zum Biegen)“. Die Konventionen nennen diesen Fall jedoch nicht. Er ist bzgl. der computergestützten Verarbeitung weniger problematisch als der umgekehrte Fall. Wenn er zugelassen werden soll, sollte er auch in den Konventionen behandelt werden.

1.2. Visualisierung

Verbaltranskriptionen können wie gewohnt als Transkripte in Zeilennotation visualisiert werden, wobei eine Zeileneinheit entweder einem Turn oder einem Ereignis, das keinem Sprecher zugeordnet ist, entspricht. Umgekehrt können Transkriptionen in Zeilennotation auch in dieser Schreibweise in einem Texteditor eingegeben werden und sollten sich dann zuverlässig parsen lassen. Da die explizite Markierung von Überlappungen auf dieser Stufe nicht gefordert wird, werden allerdings Überlappungen bei dieser Schreibweise i.d.R. nicht visualisiert werden können. Gleichfalls lassen sich Verbaltranskriptionen in Partiturnotation visualisieren und können auch in Werkzeugen, die mit dieser oder einer verwandten Notationsform arbeiten, eingegeben werden.

- Bei der Visualisierung werden Turns und andere Ereignis grundsätzlich aufsteigend nach ihrem Startpunkt geordnet. Die Konventionen machen derzeit jedoch keine eindeutige Angabe dazu, welche Ordnung bei Turns mit identischem Startpunkt zu wählen ist. Das kanonische Vorgehen scheint in diesem Falle eine nach Endpunkt absteigende Ordnung zu sein, also später endende (bzw. längere) Turns zuerst zu visualisieren. Dies könnte in den Konventionen kurz angesprochen werden.
- Wie bereits oben angesprochen, wäre zu überlegen, ob und in welcher Weise absolute Zeitwerte in die Visualisierung der Transkription eingehen sollen. Für die Analyse oder eine Veröffentlichung sind sie u.U. eher störend und sollten weggelassen werden. Wenn ein Texteditor zur Eingabe verwendet wird, muss aber spezifiziert werden, in welcher Form Transkribenden absolute Zeitwerte dort vermerken können.

2. Minimaltranskription

Gemäß einem Vorschlag auf dem GAT-Treffen im Oktober 2007 definiert die Minimaltranskription eine Stufe unterhalb der Basistranskription, in der Phrasierungseinheiten zwar markiert, aber nicht bzgl. ihrer Tonhöhenbewegung am Einheitenende klassifiziert werden. Außerdem wird auf dieser Stufe die zeitliche Ordnung durch die Markierung von Überlappungen präzisiert. Beides führt dazu, dass die Datenstrukturen deutlich komplexer werden als bei der Verbaltranskription. Es bliebe zu prüfen, ob sie sich dennoch mit allen gängigen Transkriptionswerkzeugen erzeugen und verarbeiten lassen, oder ob gewisse Werkzeuge (z.B. CLAN) dabei an ihre Grenzen stoßen.

Da das Konzept der Phrasierungseinheit bereits ein spezifisch sprachanalytisches ist, wird auf dieser Stufe auch ein gutes Maß an Unabhängigkeit gegenüber Theorie und Untersuchungszweck preisgegeben. Allerdings sehen viele andere Transkriptionssysteme eine Segmentierung vor, die strukturell der Unterteilung in Phrasierungseinheiten vergleichbar ist. Insofern bleiben die bei der Minimaltranskription entstehenden Datenstrukturen prinzipiell auch für andere Herangehensweisen nutzbar.

2.1. Modell

2.1.1. Phrasierungseinheiten

Neben der zeitlichen Ordnung sind Phrasierungseinheiten der Bereich der Konventionen, in dem hinsichtlich einer computertechnischen Umsetzung die meisten Unklarheiten bestehen.

Dabei geht es erstens um die Definition dieser Einheit selbst, insb. um klare Kriterien, die einem Transkribenden ermöglichen, Phrasierungseinheiten als solche zu identifizieren, und die sicherstellen, dass die Unterteilung der Transkription in Phrasierungseinheiten intersubjektiv verlässlich vorgenommen werden kann. Da dies Thema einer eigenen Arbeitsgruppe ist, soll hier nicht näher darauf eingegangen werden.

Zweitens geht es um Fragen der Darstellung in den Konventionen:

- Es werden alternierend die Begriffe ‚Phrasierungseinheit‘, ‚Intonationseinheit‘, ‚prosodische Einheit‘ und ‚Konstruktionseinheit‘ verwendet, offenbar immer um ein und dieselbe Einheit zu benennen. Besser wäre es, beim ersten Auftreten des Begriffs seine möglichen Alternativen zu nennen und im Folgenden bei einer Bezeichnung zu bleiben.
- Der Unterschied zwischen ‚Turn‘ und ‚Phrasierungseinheit‘ muss klarer herausgestellt werden. Es werden an mehreren Stellen Formulierungen wie „Turns oder Einheiten“ verwendet, die potentiell verwirrend sind.
- Gleichfalls wird an mehreren Stellen davon gesprochen, dass gewisse Fragen, die mit der Segmentierung in Phrasierungseinheiten zu tun haben, „theoretisch umstritten“ seien. Es wäre in diesen Fällen hilfreich, wenn der Transkribend die widerstreitenden Positionen erläutert bekäme und ihm eine Entscheidungshilfe an die Hand gegeben würde, welche Position er im Einzelfall bevorzugt einzunehmen hat.

Drittens geht es darum, wie sich die Phrasierungseinheit systematisch in Beziehung setzen lässt zur Einheit „Turn“ und zu den Grundbausteinen der Transkription. Es scheint zumindest klar, dass Turns und Phrasierungseinheiten ineinander eingebettete Einheiten sind, d.h. jede Phrasierungseinheit ist Bestandteil genau eines Turns. Wahrscheinlich ist es auch der Fall, dass ein Grundbaustein (also ein Wort, eine Pause etc.) nie Bestandteil von mehr als einer Phrasierungseinheit ist. Unklar ist jedoch:

- Ist die Unterteilung von Turns in Phrasierungseinheiten exhaustiv? D.h.: Kann/soll jeder Turn vollständig in Phrasierungseinheiten zerlegt werden? Oder können bei der Zerlegung Wörter,

Pausen, nonverbale Handlungen etc. übrig bleiben, die zwar Bestandteil des Turns, nicht aber Bestandteil einer Phrasierungseinheit sind? Wenn ja, wie sind diese übrig gebliebenen Bestandteile zu bezeichnen?

- Was genau ist der Status von unakzentuierten Anhängseln? Sind diese Phrasierungseinheiten, sind sie keine, oder sind sie Phrasierungseinheiten eines besonderen Typs? Wenn sie keine Phrasierungseinheiten sind, als was sind sie zu bezeichnen?

2.1.2. Silben

Die Einheit der Silbe ergibt sich nur bei der Annotation von Primärakzenten. Silben in Wörtern, die keinen Primärakzent tragen, werden also auch nicht als solche gekennzeichnet, die Einheit wird nicht exhaustiv erfasst. Es gibt lediglich ein Detailproblem:

- In aller Regel ist die Einheit der Silbe der Einheit des Wortes hierarchisch untergeordnet. Theoretisch ist es jedoch möglich, dass eine Silbe zwei transkribierte Wörter umfasst, nämlich bei klitischen Formen wie ‚ich HAB S doch gewusst‘. Können solche Fälle ausgeschlossen werden?

2.1.3. Zeitliche Ordnung

Auf der Stufe der Minimaltranskription wird die zeitliche Ordnung verfeinert, indem die Start- und Endpunkte überlappender Passagen verschiedener Sprecher gekennzeichnet werden. Konkret bedeutet dies, dass überall dort, wo der Turn eines Sprechers beginnt bzw. endet, der korrespondierende Punkt in bereits begonnenen bzw. noch nicht beendeten Turns anderer Sprecher identifiziert werden muss. Beispielsweise sieht eine Simultanpassage nach der Verbaltranskription folgendermaßen aus:

A: [1]ich rede viel es ist mir egal ob mich jemand unterbricht[4]
B: [2]ich möchte aber auch mal was sagen[3]

Dabei ist bereits festgelegt, dass der Turn von Sprecher A vor dem Turn von Sprecher B beginnt und nach diesem endet. Die genaue Markierung der Überlappung für die Minimaltranskription bedeutet in diesem Fall, dass die Stellen im Turn von A identifiziert werden, an denen der Turn von B beginnt und endet, etwa:

A: [1]ich rede [2]viel es ist mir egal ob[3] mich jemand unterbricht[4]
B: [2]ich möchte aber auch mal was sagen[3]

Dieses Prinzip ist auch anwendbar, wenn es um Simultanpassagen in Turns von mehr als zwei Sprechern geht. Dabei ist es für Zwecke der computertechnischen Umsetzung unbedingt notwendig, dass die Identifizierung des Endpunktes einer Überlappung nicht – wie in der derzeitigen Fassung der Konventionen – optional bleibt.

Allerdings können nach der Verbaltranskription neben den Start- und Endpunkten von Turns bereits weitere (absolute, turninterne) Zeitpunkte vorliegen, die in diesem Schritt ebenfalls in allen evtl. parallel verlaufenden Turns anderer Sprecher eingetragen werden müssten. In diesem Zusammenhang wäre weiterhin zu klären, ob auch die (zwischen Start- und Endpunkt des jeweiligen Turns liegenden) Start- und Endpunkte von Phrasierungseinheiten in allen Turns vorhanden sein müssen. Für die gewohnte Darstellung der Transkription in Zeilennotation wäre das eine unabdingbare Voraussetzung. Im obigen Beispiel könnte dies in etwa folgendermaßen aussehen (dabei ist [2a] der Startpunkt einer Phrasierungseinheit von Sprecher A, zu dem der korrespondierende Zeitpunkt im Turn von Sprecher B gefunden werden muss):

A: [1]ich rede [2]viel [2a]es ist mir egal ob[3] mich jemand unterbricht[4]
B: [2]ich möchte [2a]aber auch mal was sagen[3]

Bislang ist außerdem nicht geregelt, wie Einsatzpunkte innerhalb von Einheiten, die nicht nach dem phonographischen Prinzip beschrieben sind (also Nonverbales, Unverständliches und Pausen), zu markieren sind. Eine Möglichkeit wäre, solche Einheiten am betreffenden Punkt zu teilen und statt einer zwei äquivalente Beschreibungen zu verwenden. Bei Pausen müsste demnach deren Dauer auf zwei Elemente aufgeteilt werden, die Beschreibung von Nonverbalem müsste verdoppelt werden, also z.B.:

A: [1]erst (1.2) Pause dann ((fuchtel)) Geste[4]
B: [2]was du nicht sagst[3]

A: [1]erst (0.7) [2](0.5) Pause dann ((fuchtel))[3] ((fuchtel)) Geste[4]
B: [2]was du nicht sagst[3]

2.1.4. Nichtselbstständige Einheiten

Haupt-/ Primärakzent

An nicht selbstständigen Einheiten kommt auf der Stufe der Minimaltranskription lediglich der Haupt- bzw. Primärakzent hinzu, weil dieser mit entscheidend für die Bestimmung der Grenzen von Phrasierungseinheiten ist. Bezugseinheit des Haupt-/Primärakzents ist die Silbe, siehe dazu oben.

2.2. Visualisierung

Die Visualisierung für die Ausgabe kann, sofern die Unklarheiten bzgl. der Phrasierungseinheiten beseitigt sind, auch für die Minimaltranskription in gewohnter Weise in Zeilennotation erfolgen, d.h. jede Phrasierungseinheit (bzw. deren Komplementäreinheit(en), wenn solche festgelegt wurden) steht dann auf einer eigenen Zeile, Start- und Endpunkte von Überlappungen werden durch öffnende und schließende eckige Klammern repräsentiert.

- In diesem Zusammenhang sei noch einmal daran erinnert, dass der Status von unakzentuierten Anhängseln geklärt werden muss. Wenn diese auf Modellebene als normale Phrasierungseinheiten behandelt werden, gibt es keine Möglichkeit, sie auf der Visualisierungsebene anders (nämlich nicht als eigenständige Zeilen) zu behandeln.

Größere Änderungen sind aber notwendig, um auch weiterhin die Eingabe in einem Texteditor zu ermöglichen:

1. Die Praxis, das Ende von Phrasierungseinheiten bei der Eingabe durch ein Zeilenende zu markieren, ist problematisch, da sich so fortwährende Änderungen der zeitlichen Ordnung von Zeilen ergeben, die der Transkribend jeweils „mitdenken“ müsste, und die eine aufwändige Neuordnung der Transkriptzeilen erfordern können. Besser wäre es, wie beim GAT-Treffen im Oktober bereits vorgeschlagen, die Grenzen zunächst durch den Einsatz eines geeigneten Zeichens (z.B. den Schrägstrich) kenntlich zu machen und die Einheit der Zeile weiterhin für den gesamten Turn zu reservieren, also...

A: ich rede viel/es ist mir egal ob mich jemand unterbricht
B: ich möchte aber auch mal was sagen

... statt ...

A: ich rede viel
B: ich möchte aber auch mal was sagen
A: es ist mir egal ob mich jemand unterbricht

Diese Vorgehensweise hätte darüber hinaus den Vorteil, dass sie auch in Transkriptionssinstrumenten, die kein an die Zeilennotation angelehntes Interface haben, unverändert anwendbar bliebe.

2. Damit überlappende Passagen auch bei der Eingabe in einem Texteditor eindeutig interpretierbar sind, müsste die gängige Praxis dort verändert werden, z.B. indem die betreffenden Klammern indiziert werden, für obiges Beispiel also:

A: ich rede [viel]1/[es ist mir egal ob]2 mich jemand unterbricht
B: [ich möchte]1 [aber auch mal was sagen]2

Es sei hier noch einmal darauf hingewiesen, dass das Markieren von Überlappungen in Zeilennotation erfahrungsgemäß bei Gesprächen mit mehreren Teilnehmern sehr unüberschaubar werden kann und viele Transkribenden überfordert. Die Konventionen könnten darauf hinweisen, dass die Markierung von Überlappungen mit geeigneten spezialisierten Werkzeugen i.d.R. wesentlich einfacher und zuverlässiger durchgeführt werden kann als in einem Texteditor.

Die Visualisierung einer Silbe mit Primärakzent erfolgt durch Großschreibung. Hier müsste noch eine Detailfrage geklärt werden:

- Für das ‚ß‘ gibt es keinen korrespondierenden Großbuchstaben. Wie wäre etwa eine erste betonte Silbe im Wort ‚Fußball‘ zu visualisieren? Als ‚FUSball‘, ‚FUSSball‘ oder FUßball‘?

3. Basistranskription

Auf der Stufe der Basistranskription werden vier weitere Annotationen eingeführt, nämlich die Tonhöhenbewegung am Einheitenende, der schnelle Anschluss, die Dehnung und der interpretierende Kommentar. Die beiden letzteren erhöhen noch einmal die Komplexität der Datenstruktur.

3.1. Modell

3.1.1. Phoneme

Die Bezugseinheit für die Annotation einer Dehnung ist in der Regel ein einzelnes Phonem. Allerdings werden einzelne Phoneme teilweise als mehrteilige Grapheme repräsentiert (etwa bei ‚sch‘). Wie die Einheit der Silbe, wird auch die Einheit des Phonems nur dort gekennzeichnet, wo sie für die betreffende Annotation benötigt wird, d.h. auch die Unterteilung von Wörtern (bzw. Silben) in Phoneme erfolgt nicht exhaustiv.

3.1.2. Nichtselbstständige Einheiten

Tonhöhenbewegung am Einheitenende

Hier werden fünf Werte – ‚hoch steigend‘, ‚mittel steigend‘, ‚gleich bleibend‘, ‚mittel fallend‘ und ‚tief fallend‘ – unterschieden. Bezugseinheit ist die Phrasierungseinheit bzw. auch unakzentuierte Anhängsel. Da die Schwierigkeiten bzgl. der Segmentierung in Phrasierungseinheiten bereits auf der Ebene der Minimaltranskription behandelt werden, sollten durch diese Annotation keine neuen Probleme entstehen.

Schneller Anschluss („Latching“), Verschleifung

Ein schneller Anschluss kann zwischen zwei unmittelbar nacheinander geäußerten Wörtern eines Sprechers bestehen. Die Konventionen schreiben vor, schnelle Anschlüsse im Normalfall nur an der Grenze einer Phrasierungseinheit zu annotieren, innerhalb der Phrasierungseinheit aber sog. „Verschleifungen“ mit dem gleichen Zeichen kenntlich zu machen (S. 9). Als Bezugseinheit können in beiden Fällen entweder beide beteiligten Wörter oder nur das zweite von diesen angenommen werden. Letzteres wäre hinsichtlich der Datenstruktur einfacher zu handhaben und bringt auch sonst wahrscheinlich keine Probleme mit sich.

- Es ist nicht ganz klar, ob der schnelle Anschluss einer Phrasierungseinheit und die Verschleifung zweier Wörter innerhalb einer Phrasierungseinheit als das gleiche oder als verschiedene Phänomene anzusehen sind. In einer Fußnote (S.9) heißt es: „Die Mehrdeutigkeit des Zeichens „=“ für Latching und Verschleifung wird also durch die Stelle, an der es notiert wird, disambiguiert.“ Demnach handelt es sich also eher um verschiedene Phänomene. Die Desambiguierung anhand der Position kann auch beim computergestützten Parsing der Transkription angewendet werden. Einfacher und klarer wäre es jedoch, für die unterschiedlichen Phänomene auch unterschiedliche Zeichen zu verwenden.

Dehnung

Hier werden nach Länge der Dehnung drei Werte unterschieden. Bezugseinheit ist das Phonem (s.o.).

- Die Vorgabe, wie die Länge der Dehnung bestimmt werden soll (S.11), ist nicht klar. Die Formulierungen „als Faustregel kann vielleicht gelten“, „etwas länger“ und „deutlich länger“ sowie die Anführungsstriche um das Wort „normal“ ergeben zusammen genommen keine Anweisung, die ein Transkribend zuverlässig befolgen könnte.

Interpretierende Kommentare

Interpretierende Kommentare nehmen die Form einer freien Beschreibung (z.B. ‚empört‘, ‚erstaunt‘) an. Sie werden formal nicht von Beschreibungen von nonverbalen Handlungen, die parallel zu Verbalem stattfinden (z.B. ‚weinend‘, ‚hustend‘), unterschieden. Eine eindeutige Bezugseinheit ist nicht festzustellen – interpretierende Kommentare können sich im Prinzip auf beliebige Abschnitte der transkribierten Rede eines Sprechers beziehen. Damit können sie auch potentiell die Hierarchie Turn – Phrasierungseinheit – Wort – Silbe – Phonem durchbrechen, z.B. wenn eine parallele nonverbale Handlung sich auf die zweite Hälfte einer Phrasierungseinheit und die erste Hälfte der folgenden bezieht...

A: ich will das <<weinend> nicht mehr hören
B: aber ich höre jetzt auf> zu weinen

... oder wenn ein interpretierender Kommentar innerhalb eines Wortes beginnt und sich über weitere Wörter erstreckt:

A: sind sie viel<<zögernd>leicht von der polizei>

Da dieser Umstand weit reichende Konsequenzen für die Datenstruktur hat, wäre es wünschenswert, ihn etwas einzuschränken, z.B. indem gefordert wird, dass interpretierende Kommentare nicht über Turngrenzen hinweg gelten sollen. Ggf. müssten in einem solchen Fall dann zwei identische Kommentare verwendet werden.

A: mein gott <<hustend> dieser schreckliche husten
den werde ich so schnell wohl nicht wieder los>
B: was sagst du
A: <<hustend> ich sagte dieser schreckliche husten>

3.2. Visualisierung

Bzgl. der Visualisierung für die Ausgabe ergeben sich keine grundsätzlich neuen Probleme. Für die Eingabe in einem Texteditor müsste Folgendes überlegt werden:

- Wie soll die Annotation der Tonhöhenbewegung am Einheitenende vorgenommen werden? Sollen die nach der Verbaltranskription vorhandenen Schrägstriche, die das Ende einer Phrasierungseinheit kenntlich machen, durch Fragezeichen, Komma, Punkt etc. ersetzt werden? Es sei noch einmal daran erinnert, dass der Punkt zu diesem Zwecke von seinen weiteren Bedeutungen „befreit“ werden müsste.
- Wie soll das Latching markiert werden? Am einfachsten zu handhaben wäre es, wenn auch hier einfach das zweite betroffene Wort markiert, also mit einem Gleichheitszeichen am Anfang versehen würde.
- Wenn die Markierung einer Dehnung sich auf ein mehrteiliges Graphem bezieht, ist es nicht trivial, dies in die entsprechende Datenstruktur zu übersetzen. Kann man diese Fälle systematisch auflisten und Mehrdeutigkeiten dabei ausschließen?

4. Feintranskription

Auf der Stufe der Feintranskription kommen zahlreiche nicht-selbstständige Einheiten hinzu, die vor allem prosodische Eigenschaften der segmentalen Transkriptionseinheiten beschreiben. Da auf dieser Stufe weit reichende Änderungen im Zuge der GAT-Überarbeitung zu erwarten sind, sind die Ausführungen hier etwa kürzer gehalten.

4.1. Modell

4.1.1. Nichtselbstständige Einheiten gemäß GAT 1

Neben-/Sekundärakzent und extra starker Akzent

Bezugseinheit dieser beiden Annotationen ist wie beim Hauptakzent jeweils die Silbe. Dabei wird der Nebenakzent zusätzlich zum Hauptakzent markiert, während der extra starke Akzent sich in der Regel auf Silben beziehen wird, die bereits als betont gekennzeichnet sind.

Auffällige Tonhöhen sprünge

Hier werden vier Werte – ‚auffälliger größerer‘ bzw. ‚kleinerer‘ Tonhöhen sprung ‚nach oben‘ bzw. ‚nach unten‘ – unterschieden. Als Bezugseinheit sollte die nächste folgende Silbe angenommen werden.

Verändertes Tonhöhenregister

Hier werden zwei Werte – ‚tiefes‘ und ‚hohes‘ Register – unterschieden. Laut Konventionstext (S. 18) werden diese „für eine oder mehrere Phrasierungseinheiten oder für Teile eine Einheit“ annotiert. Die Bezugseinheit ist also – wie beim interpretierenden Kommentar – ein prinzipiell beliebiger Abschnitt eines Turns.

Akzenttonhöhenbewegungen

Hier werden fünf mögliche Werte unterschieden: fallend, steigend, gleich bleibend, steigend-fallend und fallend-steigend. Die Annotation dieser Einheit „bezieht sich auf die Tonhöhenbewegung in und nach der Akzentsilbe“ (S. 18), die damit als Bezugseinheit gelten kann.

Lautstärke- und Sprechgeschwindigkeitsveränderungen

Bei der Lautstärke werden sechs Werte (forte, fortissimo, piano, pianissimo, crescendo und diminuendo), bei der Sprechgeschwindigkeit vier Werte (allegro, lento, accelerando, rallentando) unterschieden. Auch hier ist die Bezugseinheit ein beliebiger Abschnitt eines Turns.

4.1.2. Nichtselbstständige Einheiten gemäß GAT 2

Beim GAT-Treffen im Oktober wurden weitere Einheiten genannt, die für die Stufe der Feintranskription eingeführt werden könnten, z.B. die „Knarrstimme“. Solche Einheiten müssten, wenn ihre Hinzunahme gewünscht ist, ebenfalls bezgl. ihrer möglichen Werte und ihrer Bezugseinheiten analysiert werden.

4.2. Visualisierung

Tonhöhenregister sowie Lautstärke- und Sprechgeschwindigkeitsveränderungen werden analog zu interpretierenden Kommentaren visualisiert, Nebenakzent und extra starker Akzent analog zum Hauptakzent. Für die übrigen Einheiten werden an der betreffenden Stelle Zeichen in den Text eingefügt. Hier ist es unbedingt notwendig, diese Zeichen genauer zu spezifizieren:

- Was ist der Unterschied zwischen den Zeichen für kleinere und größere Tonhöhensprünge? Sind dies wirklich distinkte Zeichen oder handelt es sich um eine unterschiedliche Formatierung ein und desselben Zeichens?
- Die Zeichen, die zum Markieren von Akzenttonhöhenbewegungen verwendet werden, sind in dreierlei Hinsicht problematisch. Erstens bestehen bei allen zahlreiche Verwechslungsmöglichkeiten (z.B. zwischen Akut/Gravis-Akzent und Apostroph oder einfachem Anführungszeichen). Zweitens ist nicht garantiert, dass sie in jedem gängigen Zeichensatz enthalten sind. Drittens lassen sie sich nicht in jeder Konfiguration ohne weiteres über die Tastatur eingeben. Hier müsste unbedingt mehr Klarheit geschaffen werden.

Zusammenfassung

Dieser Text hat verschiedene Änderungen der derzeitigen Fassung der GAT-Konventionen vorgeschlagen, die die Eignung von GAT für computergestützte Verfahren verbessern würden. Ein großer Teil davon betrifft Detailänderungen im Gebrauch von Transkriptionszeichen und in den Formulierungen des Konventionstextes. Darüber hinaus werden einige grundlegendere Modifikationen vorgeschlagen, die hier abschließend noch einmal zusammengefasst seien:

1. eine verfeinerte Modularisierung der Konventionen durch eine Unterscheidung von vier statt zwei Transkriptionsstufen: Verbal-, Minimal-, Basis- und Feintranskription,
2. eine neue Vorgabe zum Einfügen von Verweisen aus der Transkription in die Aufnahme durch absolute Zeitwerte,
3. eine veränderte Vorgabe zur Repräsentation der zeitlichen Ordnung, insbesondere zum Vorgehen bei der Markierung von zeitlich parallelen Gesprächsbestandteilen,
4. eine präzisierte Definition des Begriffs „Turn“ und ggf. die Definition einer oder mehrerer komplementärer Einheiten auf der Turnebene,
5. eine präzisierte Definition des Begriffs „Phrasierungseinheit“ und ggf. die Definition einer oder mehrerer komplementärer Einheiten auf der Ebene der Phrasierungseinheit.

Mit diesen Änderungen sollte es sowohl möglich bleiben, GAT mit einem einfachen Texteditor anzuwenden, als auch spezialisierte Transkriptionssoftware zu nutzen (und ggf. zu entwickeln), die das Transkribieren mit GAT effizient unterstützt.

Literatur

Edwards, Jane (1995): Principles and alternative systems in the transcription, coding and mark-up of spoken discourse. In: Leech, Geoffrey / Myers, Greg / Thomas, Jenny (Hrsg.): Spoken English on Computer: Transcription, Markup and Application. Harlow: Longman, 19-34.

Schmidt, Thomas (2005): Computergestützte Transkription – Modellierung und Visualisierung gesprochener Sprache mit texttechnologischen Mitteln. Frankfurt a.M.: Peter Lang.

TEI-Guidelines for the Transcription of Speech. <http://www.tei-c.org>

Anhang: Schematische Übersichten

1. Transkriptionsstufen (nach Theorieabhängigkeit, Komplexität des Datenmodells)

1. Verbaltranskription
2. Minimaltranskription (Rohtranskription, Arbeitstranskription)
3. Basistranskription
4. Feintranskription

2. Einheiten

2.1. selbstständige Einheiten

Nicht-sprachliche Einheiten: Sprecher, Zeitpunkte, absolute Zeitwerte

- Ebene 2: Turns, ggf. komplementäre Einheiten
Ebene 1: Phrasierungseinheiten, ggf. komplementäre Einheiten
Ebene 0: Wörter, besondere Wörter (Abkürzungen, Zahlen, Buchstabieren), andere phonologisch beschriebene Einheiten (Lachen, Rezeptionssignale, Verzögerungssignale), unverständliche verbale Handlung, Pausen, nonverbale Handlungen und Ereignisse, Ein- und Ausatmen
Ebene -1: Silben
Ebene -2: Phoneme

- Verbaltranskription: Ebenen 0 und 2
Minimaltranskription: Ebenen -1, 0, 1 und 2
Basistranskription: alle Ebenen

2.2. Nicht-selbstständige Einheiten

2.2.1. Verbaltranskription

Alternativlautung → Wort oder n Wörter
schwere Verständlichkeit → Wort oder n Wörter

2.2.2. Minimaltranskription

Akzent → Silbe

2.2.3. Basistranskription

Tonhöhenbewegung → Phrasierungseinheit
Dehnung → Phonem
Verschleifung → Wort
interpretierender Kommentar → beliebiger Abschnitt eines Turns

2.2.4. Feintranskription:

Lautstärke → beliebiger Abschnitt eines Turns
Sprechgeschwindigkeit → beliebiger Abschnitt eines Turns
Tonhöhenregister → beliebiger Abschnitt eines Turns
Akzenttonhöhenbewegung → Silbe
Tonhöhenprung → Silbe

3. Zeichen²

3.1. Zeichen, die selbstständige Einheiten beschreiben

a-zäöüß	Phoneme
-	als Bestandteile von Bindestrichkomposita
fremdsprachliche Grapheme	z.B. in charité, à la mode, citroën
Apostroph	Glottalverschluss

3.2. Zeichenketten, die selbstständige Einheiten beschreiben

xxx	unverständliche Silbe
(.)	Mikropause
(-)	kurze Pause (<0.5 Sekunden)
(--)	mittlere Pause (<0.75 Sekunden)
(---)	längere Pause (<1.0 Sekunden)
(d+.d)	geschätzte Pause (>1.0 Sekunden)
(d+.dd)	gemessene Pause
((.+))	nonverbale Handlung, Ereignis
h, hh, hhh	Ausatmen
.h, .hh, .hhh	Einatmen
{d+.d+}	absolute Zeitangabe

3.3. Grenzmarkierer

Leerzeichen	Wortgrenze (alle Transkriptionsstufen)
/	Grenze einer Phrasierungseinheit (ab Minimaltranskription)
?,-;	Grenze einer Phrasierungseinheit (ab Basistranskription)
((Beginn einer nonverbalen Handlung, eines Ereignisses
))	Ende einer nonverbalen Handlung, eines Ereignisses
{	Beginn einer absoluten Zeitangabe
}	Ende einer absoluten Zeitangabe
(Beginn einer schwer verständlichen Passage
)	Ende einer schwer verständlichen Passage
<<.+>	Beginn eines interpretierenden Kommentars
>	Ende eines interpretierenden Kommentars
A-Z	Beginn oder Ende einer akzentuierten Silbe
!	Beginn oder Ende einer extra stark akzentuierten Silbe

² Manche Zeichen, z.B. die meisten Grenzmarkierer werden doppelt aufgeführt. Zeichenketten sind teilweise als reguläre Ausdrücke beschrieben, wobei

d = Ziffer

. = beliebiges Zeichen

+ = ein- oder mehrmaliges Auftreten

| = Alternativen

3.4. Zeichen und Zeichenketten, die zur Visualisierung von nicht-selbstständigen Einheiten verwendet werden

Verbaltranskription

(Beginn einer schwer verständlichen Passage
)	Ende einer schwer verständlichen Passage
/	trennt innerhalb schwer verständlicher Passagen verschiedene Alternativlautungen voneinander

Minimaltranskription

A-Z	akzentuierte Phoneme
-----	----------------------

Basistranskription

=	schneller Anschluss
:	Dehnung
<<.,+>	Beginn eines interpretierenden Kommentars
>	Ende eines interpretierenden Kommentars
?,-,:;	Tonhöhenbewegung am Einheitenende

Feintranskription

A-Z	akzentuierte Phoneme
!	extra stark akzentuierte Silbe
↑↓	Tonhöhen sprünge
↑↓↑↓	Akzenttonhöhenbewegung
fünf verschiedene Diakritika	Akzenttonhöhenbewegung
<<h>, <<t>	Beginn eines hohen bzw. tiefen Tonhöhenregisters
<<f>, <<ff>, etc.	Beginn einer Lautstärkeveränderung
<<all>, <<len>, etc.	Beginn einer Sprechgeschwindigkeitsveränderung
>	Ende einer Lautstärke- oder Sprechgeschwindigkeitsveränderung bzw. eines markierten Tonhöhenregisters