

How to:
Segmentierung

Dieses Dokument erläutert die Verwendung der eingebauten Segmentierungsalgorithmen der endlichen Maschinen des EXMARaLDA Partitur-Editors.

Es sind keine Vorkenntnisse über Algorithmen, endliche Maschinen oder reguläre Sprache für das Verständnis der ersten beiden Kapitel nötig, weil alle relevanten Aspekte in diesem Dokument erklärt werden. Dennoch sollten Sie vorab

- Understanding the basics of EXMARaLDA

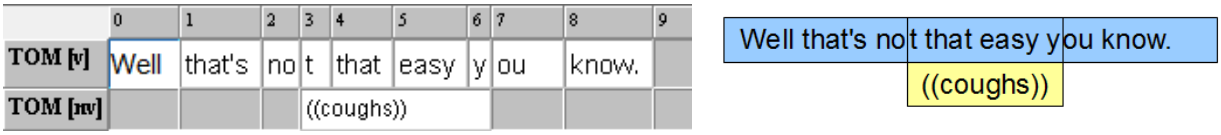
gelesen haben.

Inhalte

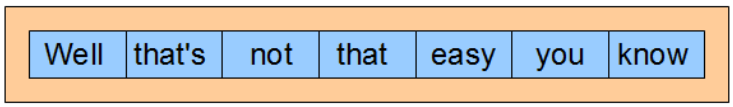
A. Über EXMARaLDA und Segmentierung

1. Basis- und segmentierte Transkriptionen

Im Grunde genommen gibt es nicht *das eine* EXMARaLDA-Transkriptionsformat, sondern zwei: Die Basistranskription und die segmentierte Transkriptions. Bei der, im Partitur-Editor erstellten, Transkription handelt es sich um eine EXMARaLDA Basistranskription (daher die Dateierweiterung .exb[1]). Obwohl die Ereignisse mit ihren Grenzen als Segmente angezeigt werden, handelt es sich hierbei lediglich um visuelle Segmente. Die Elemente auf der Zeitachse (bzw. die Zeitpunkte) dienen der Organisation verschiedener Ereignisse. Sie geben Auskunft über deren Reihenfolge, Gleichzeitigkeit und Überlappungen. Abgesehen von der organisatorischen Information über diese Ereignisse in der Basistranskription, gibt es keine sprachlichen oder anderweitig transkriptionsrelevanten Bedeutungen. Das hat zur Folge, dass die Grenzen nicht als Wortgrenzen erkannt werden, selbst wenn für jedes Einzelwort ein separates Ereignis erzeugt wird. Der Segmentierungsprozess wird immer Ereignisse als solche erkennen, aber wie Sie der Grafik entnehmen können, müssen sie Ereignisse eventuell anders gebrauchen um Überlappungen zu beschreiben, wodurch deren Bedeutung verloren gehen würde. Der Segmentierungsprozess beruht ausschließlich auf Leerzeichen und Interpunktion als Indikatoren für Wortgrenzen.



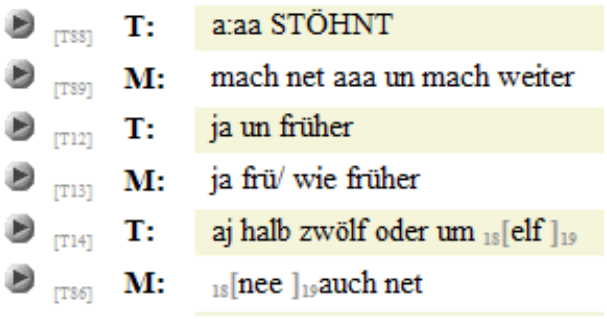
Wörter oder andere Einheiten, die durch verschiedene Symbole der jeweiligen Transkriptionskonvention gekennzeichnet werden, werden vor der Segmentierung nicht erkannt und somit kann auch die Richtigkeit der Transkription nicht ausgewertet werden.Fehlerfreie Basistranskriptionen werden durch die Segmentierung in ein anderes Format, die EXMARaLDA segmentierte Transkription, (Dateierweiterung .exs) umgewandelt. Das segmentierte Format drückt die Informationen über verschiedene Einheiten in der Transkription explizit aus und wird teilweise nach ihnen organisiert, z. B. eine aus Wörtern bestehende Äußerung.



Mit der Information über die unterschiedlichen Segmente lassen sich viele verschiedene Visualisierungen aus ein und derselben Transkription erstellen. Die Eingabe erfolgt im Partitur-Editor, der auf dem Format der musikalischen Partitur basiert. Die Ausgabe hingegen kann in einem komplett anderen Format erfolgen. Enthält die Segmentierung Ihrer Transkription Äußerungen wie z. B. in HIAT, können Sie eine zeitlich sortierte Äußerungsliste mit Markierung der Überlappungen generieren. Die Äußerungsliste entspricht einem vertikalen Transkriptionsformat und hat nur wenig Ähnlichkeit mit dem Format der musikalischen Partitur. Zum Herunterladen verschiedener Stylesheets für verschiedene Visualisierungen sei auf den Download Bereich auf www.exmaralda.org verwiesen.

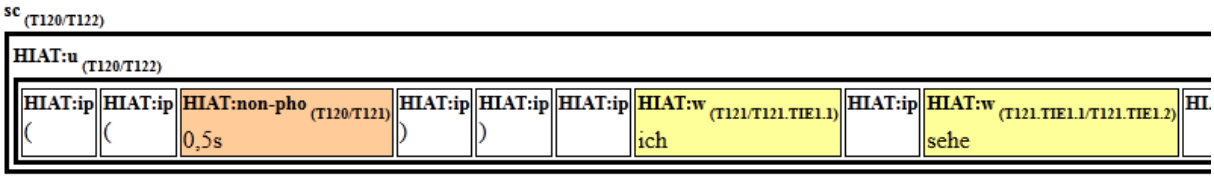
[1] Die Erweiterungen .exb und .exs sind nur für die EXMARaLDA-Instrumente von Interesse. Alle EXMARaLDA-Dateien sind XML- Dateien.

Es können auch andere Einheiten der segmentierten Transkription für weitere Visualisierungen und Darstellungsformate verwendet werden: Durch Sammeln, Sortieren und Auszählen der Wörter, kreiert EXACT Wortlisten mit Häufigkeiten für jedes Korpus mit Wortsegmentierung.



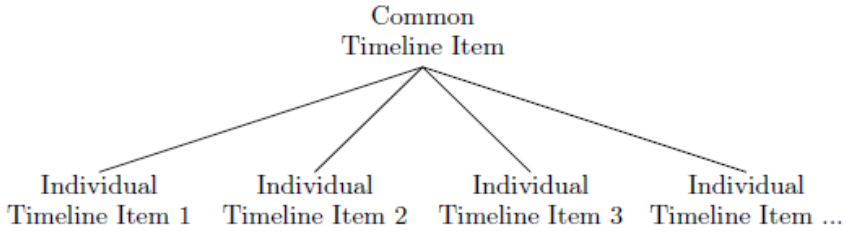
Word	Frequency
dann	833
ja	810
du	712
nach	641
und	598
äh	568
links	372
ich	368
bis	355

Segmentierte Transkriptionen können im Partitur-Editor weder geöffnet noch eingesehen werden. Abgesehen vom Arbeiten mit der XML-Datei besteht z.Z. nicht die Möglichkeit das Resultat der Segmentierung zu überprüfen oder zu ändern. Falls Sie das Resultat der Segmentierung trotzdem einsehen möchten, gibt es im Download Bereich auf www.exmaralda.org ein Stylesheet für die HTML-Visualisierung („Segmentierte Transkription als HTML-Darstellung"), das auf die segmentierte Transkription angewendet werden kann.



Ein sehr wichtiges Merkmal der segmentierten Transkription sind sie sogenannten *'Zeitgabeln'* (time forks). Da eine manuelle Alignierung parallel verlaufender Ereignisse (z.B. bei gleichzeitigem Sprechen) im Partitur-Editor nicht erforderlich ist, gibt es auch keine gemeinsame Zeitachse mit Zeitpunkten (Elementen auf der Zeitachse) für jedes neues Segment in der segmentierten Transkription. Das EXMARaLDA-Datenmodell erlaubt es, dass die Transkriptionsspuren *nicht-gemeinsame* Zeitpunkte haben. Daher werden spurspezifische oder individelle Elemente auf der Zeitachse "innerhalb" oder "unterhalb" des gemeinsamen Zeitachsen-Punktes als Zeitgabeln dargestellt. Sie sind absolut unabhängig von anderen Elementen auf der Zeitachse anderer Transkriptionsspuren.

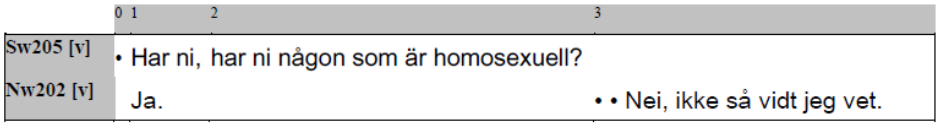
Wenn Tom „Hello Tim!" und Tim „Hi Tom, how are you?" sagt, wissen wir zwar, dass die Äußerungen zur gleichen Zeit anfangen und enden, können aber keine Informationen über Überlappungen und deren Ausmaß geben. Der gemeinsame Start- und Endzeitpunkt ist Teil der gemeinsamen Zeitachse, und für individuelle Elemente auf der Zeitachse gibt es für jedes Segment in jeder Spur eine Zeitgabel. Bedenken Sie, dass der Segmentierungsalgorithmus diese Vorgehensweise vorschreibt, sobald keine gemeinsamen Elemente auf der Zeitachse existieren.



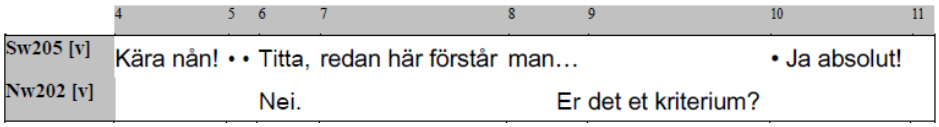
2. Segmente in Transkriptionssystemen

Transkriptionssysteme haben verschiedene Analyseeinheiten und unterschiedliche Symbole für deren Markierung. Das Transkriptionssystem HIAT nutzt beispielsweise Satzpunkte, Ausrufe- und Fragezeichen für die Markierung segmentaler Äußerungen und für die Angabe der Äußerungsmodi.

[1]



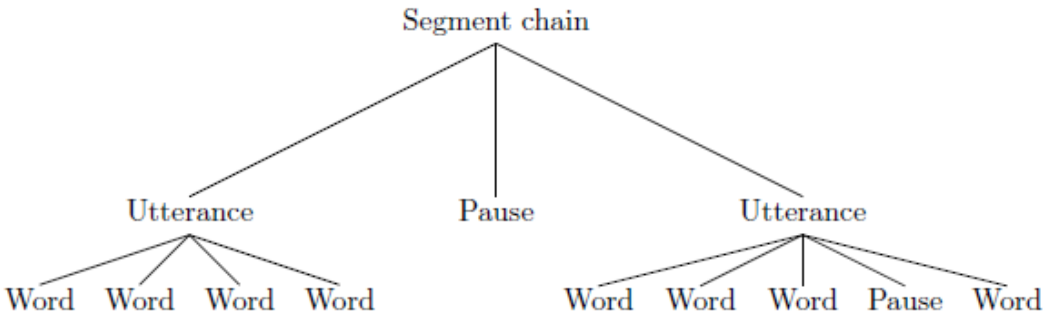
[2]



Die Kodierung über den Umfang und die Äußerungsart würde ohne die Unterstützung von HIAT im Partitur-Editor in etwa so aussehen:

Mit der eingebauten Unterstützung von Transkriptionskonventionen, lässt Sie EXMARaLDA bekannte Symbole der jeweiligen Transkription verwenden. Für alle Funktionen, die unter dem Punkt **Segmentierung** im **Transkriptionsmenü** aufgelistet sind, verwendet der Partitur-Editor dann die Symbole des Transkriptionssystems für die Segmentierung der Transkription.

Die Segmentierung kann allgemein als eine Umwandlung der Transkriptionsspuren in mehrere einfache, flache Bäume, (wie in der formellen/ generativen Grammatik) deren Strukturen vom jeweiligen Transkriptionssystem abhängen, verstanden werden:
Die größte Einheit ist *ausnahmslos* die inhärente EXMARaLDA Segmentkette, die ununterbrochene Folge von Ereignissen, die zu einem Sprecher gehören. Innerhalb einer Segmentkette kann es verschiedene Äußerungsarten geben, die wiederrum Wörter, Pausen und nichtphonologische Phänomene aufweisen können.



B. Segmentierungsalgorithmen

Der Gebrauch endlicher Maschinen ist nicht spezifisch für EXMARaLDA. Wenn Sie die Segmentierung nur für unterstützte Transkriptionssysteme nutzen möchten, können Sie den nächsten Abschnitt überspringen. Dennoch kann ein Grundverständnis über die Funktionsweise der Segmentierung helfen, Segmentierungsfehler von vornherein zu vermeiden.

1. Funktionsweise

Die Grundidee besteht darin die „Maschine" die Transkription Symbol für Symbol lesen zu lassen, um zu prüfen, ob alle Eingaben entsprechend der Konventionen getätigt worden sind und damit aus den Eingaben eine entsprechende segmentierte Transkription erstellt werden kann. Um eine korrekte Transkription erstellen zu können, dürfen in ihr nur Symbole verwendet werden, die zu den erlaubten Symbolen dieser Art von Transkription gehören. Zudem müssen sie gemäß der Regeln des Lexikons und der Syntax korrekt verwendet werden. Die Maschine muss demnach in der Lage sein, alle vorstellbaren Transkriptionen entsprechend der Transkriptionskonventionen erkennen zu können, ähnlich wie bei den natürlichen (menschlichen) Sprachen.

Glücklicherweise lassen sich Transkriptionen weniger komplex beschreiben als natürliche Sprachen und daher können die sogenannten *endlichen Maschinen* diese erkennen und produzieren. Auf der Suche nach Segmentierungsfehlern startet die Maschine bei der ersten Segmentkette der Transkriptionsspur und arbeitet sich entweder bis an das Ende der Transkription vor, oder stoppt an einer ungültigen Sequenz. Die Maschine kann sich nicht jedes gelesene Symbol merken und erkennt demnach nur den „state" des Fehlers, d.h., ob er sich „innerhalb eines Wortes einer Äußerung in einer Segmentkette" oder „in einem nicht phonologischen Phänomen innerhalb einer Äußerung in einer Segmentkette" befindet. In Abhängigkeit von dem ursprünglichen „state" und dem Input, bewegt sich die Maschine in einen anderen „state" ,d.h. sie liest Leerzeichen und ändert eine Einheit, die den Status eines Wortes hat, in etwas, das den Status einer direkten Äußerung hat.

2. Auswirkungen auf die Transkription

Da jedes Symbol, oder jede Kombination von Symbolen, ungeachtet des Kontextes als Teil der Transkriptionskonvention gelesen wird, ist die Bedeutung der Symbole nicht so flexibel wie es die menschliche Interpretation erlauben würde. Ein Beispiel: Um das Ende eines Deklarativsatzes zu markieren, verwendet man nach HIAT einen Satzpunkt und deshalb ist es nicht möglich ihn in solchen Transkriptionskonventionen als Abkürzungssymbol zu verwenden. Der Segmentierungsalgorithmus des Partitur-Editors liest die Zeichen ohne den sprachlichen Kontext einzubeziehen. Sprachspezifische Abkürzungen wie in dem Satz „Das sind z.B. Hühner" würden demnach in drei Äußerungen segmentiert werden:

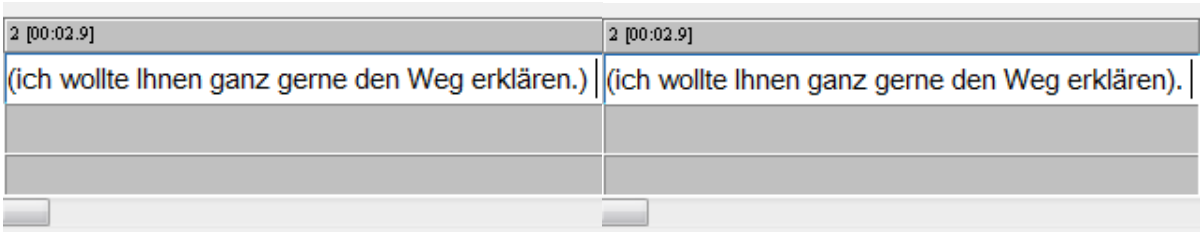
1. Das sind z.
2. B.
3. Hühner.

Für die Markierung eines abgebrochenen Satzes ist die Unterscheidung zwischen drei aneinandergereihten Satzpunkten (…) und dem Ellipsen Symbol (…) wichtig. Würden Sie die Satzpunkte aneinanderreihen, kreieren Sie damit lediglich leere, deklarative Äußerungen. Für den Leser ist die Unterscheidung klar, für den Segmentierungsalgorithmus hingegen nicht.

Einige Details, die als unwichtig erscheinen mögen, sind für die Funktionsweise von Segmentierungsalgorithmen von Bedeutung.

Ein Beispiel: Sie verwenden HIAT um eine Äußerung mit zwei undeutlichen Wörtern am Ende zu transkribieren. Dafür verwenden Sie einfache Klammern. Hierbei ist wichtig, dass die geschlossene Klammer vor dem Äußerungsendzeichen steht. Schließen Sie zuerst den unverständlichen Part *innerhalb* der Äußerung, bevor Sie die komplette Äußerung abschließen.

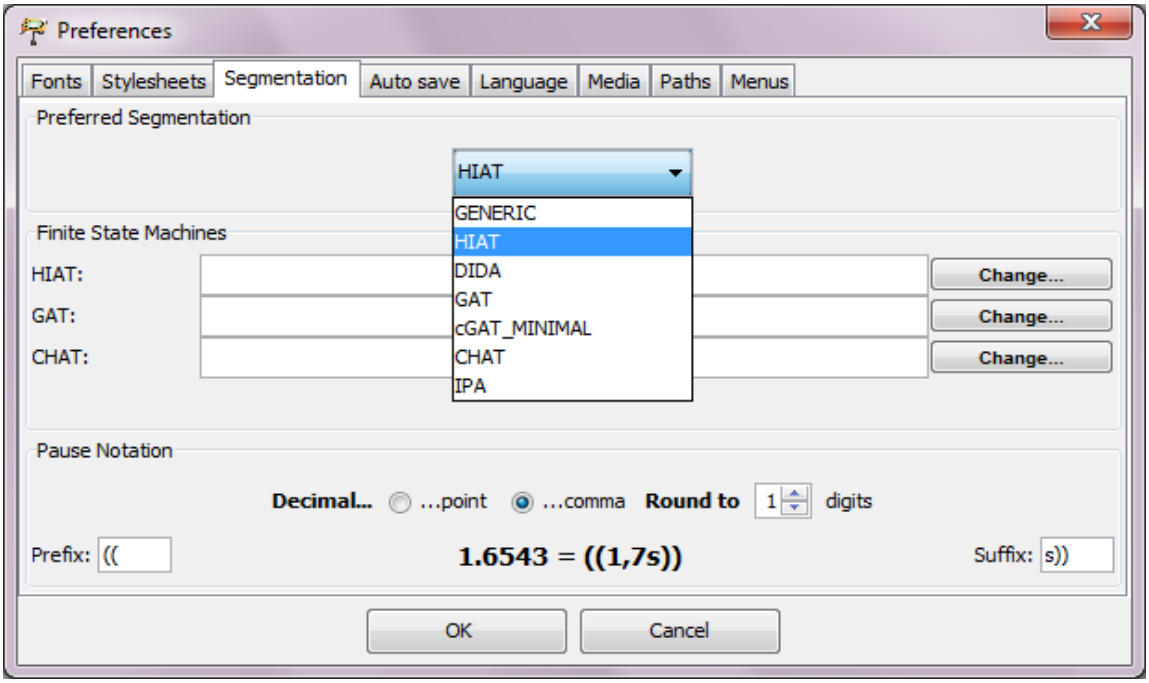
Das gleiche Vorgehen wird bei einer Äußerung gebraucht, die im Ganzen unverständlich ist, da die Äußerung vor dem unverständlichen Teil in Klammern beginnt. Bei der linken Option würden sich Segmentierungsfehler ergeben, bei der rechten Option hingegen nicht.



C. Segmentierung im Partitur-Editor

1. Segmentierungsoptionen

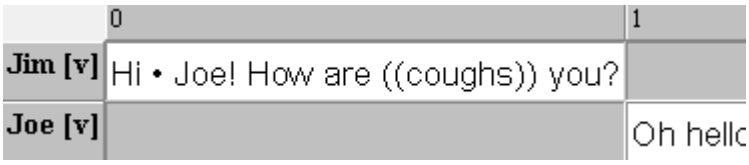
Der EXMARaLDA Partitur-Editor beinhaltet Segmentierungen für die meistgenutzten Transkriptionskonventionen (HIAT, DIDA, GAT, CHAT) und IPA. Zudem gibt es die generische Segmentierung, die für Transkriptionen ohne lineare Symbole (wie Äußerungsendzeichen) und für schriftliche Daten genutzt werden kann. Für die Beschreibung der Einheiten können Sie stattdessen mit Annotationen arbeiten. Ihre bevorzugte Segmentierung wählen Sie im Dropdown-Menü unter **Bearbeiten > Voreinstellungen... > Segmentierung** aus. Alle Segmentierungen nutzen die Notation in Segmentketten[2] als Toplevel, d.h. wenn eine Segmentkette irrtümlich unterbrochen wird, wird durch das gesplittete Segment auch die Segmentierung fehlerhaft sein.



[2] Ununterbrochene Folgen von Ereignissen, die zu einem Sprecher gehören

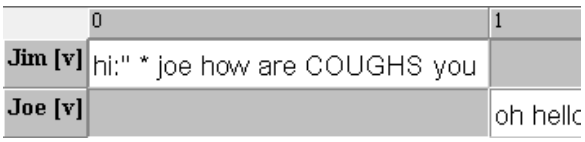
Generisch
Die generische Segmentierung kann immer verwendet werden. Sie erkennt Wörter, die durch Leerzeichen oder Nicht-Wortzeichen (Interpunktionszeichen etc.) getrennt sind, d.h. wenn keine linguistische Information gebraucht wird, und Segmentketten. Mit dieser Form der Segmentierung können Sie Wortlisten oder Segmentkettenlisten erstellen.

HIAT
Die Segmentierung nach HIAT erkennt Äußerungen, Wörter, Pausensymbole und nicht-phonologische Segmente entsprechend der HIAT-Konventionen. Andere bedeutungstragende Interpunktionszeichen wie Anführungszeichen für Zitate, Schrägstriche für Reparaturen, Bindestriche für Wortfragmente und Kommata für Wiederholungen, werden zwar erkannt, aber in der segmentierten Transkription nur als Interpunktionszeichen (IP) transferiert und nicht wie Wörter oder Äußerungen gewertet. Das bedeutet, dass es keine „Zitat-“ oder „Abbruch“ Segmente in der segmentierten Transkription gibt. Im Folgenden finden Sie die Segmentierung einer Segmentkette, die zwei Äußerungen beinhaltet, von denen die eine eine interne Pause und die andere ein internes nicht-phonologisches Ereignis beinhaltet, nämlich das Husten des Sprechers:



Segment Chain													
Utterance						Utterance							
W[3]	IP	Non-Pho	IP	W	IP	W	IP	W	IP	Non-Pho	IP	W	IP
Hi		•		Joe	!	How		are	((	coughs	))	you	?

DIDA
Die Segmentierung nach DIDA segmentiert Wörter, Pausen und nicht-morphemische Äußerungen. Sie beinhaltet auch noch andere Symbole, die in der Transkription vorkommen. Dennoch gibt es in DIDA keine Äußerungen oder gleichwertige „Brocken“. Im Folgenden sehen Sie nochmal eine Segmentierung einer Segmentkette, dieses Mal mit einer zusätzlichen Dehnung und Betonung des ersten Wortes:

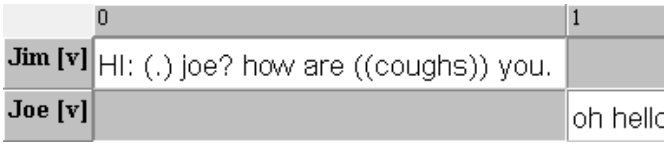


Segment Chain												
W	IP	PAUSE	IP	W	IP	W	IP	W	IP	NMÄ	IP	W
Hi:”		*		joe		how		are		COUGHS		you

[3] = Wort

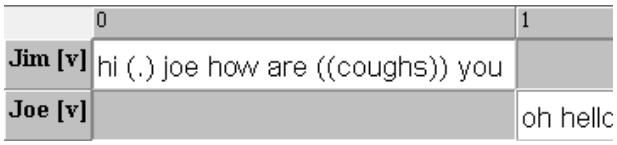
GAT und cGAT Minimal

Bei der traditionellen GAT-Transkriptionskonvention wird, auf Basis von Äußerungsendzeichen, nach Phrasierungseinheiten segmentiert. Die Begrüßung von Jim bestünde demnach lediglich aus zwei Segmenten innerhalb der Segmentkette:



Segment Chain	
PE	PE
HI: (.) joe?	how are ((coughs)) you.

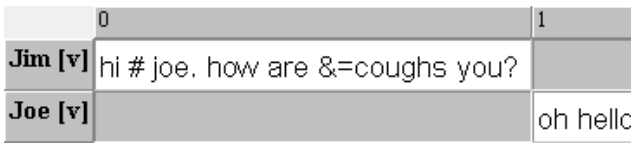
Für eine detaillierte Segmentierung wurde ein neues Set von Konventionen entwickelt: die cGAT Minimal[4] Konventionen. Die cGAT-Konventionen sind im Folker- Transkriptionshandbuch[5] in deutscher Sprache detailliert beschrieben. Eine englische Übersetzung ist in Planung. Eine kurze Beschreibung in Deutsch finden Sie im Anhang dieses Dokuments. Folgen Sie dieser Beschreibung und Sie werden in der Lage sein, Segmente und Segmentketten noch genauer segmentieren. Auf der anderen Seite unterstützen die cGAT Minimalkonventionen keine Endzeichen für Intonationphrasen. Das bedeutet, dass es z.Z. noch keine Möglichkeit gibt, die Vorteile beider Segmentierungsoptionen nutzen zu können:



Segment Chain													
W	S[6]	Pause	S	W	S	W	S	W	S	Non-Pho	S	W	S
hi		(.)		joe		how		are		((coughs))		you	

CHAT

Transkriptionen, die dem CHAT-Format folgen, welches im CHAT manual[7] beschrieben wird, werden nach Äußerungen entsprechend der Äußerungsterminatoren segmentiert. Betrachtet man eine ähnliche Segmentkette wie im vorigen Abschnitt, erhalten wir eine Segmentierung in zwei Äußerungen:



Segment Chain	
U[8]	U
Hi # joe.	how are &=coughs you?

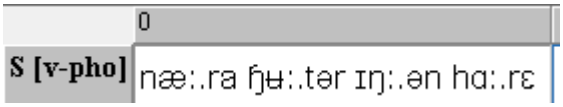
[4] http://agd.ids-mannheim.de/html/FOLKER-Transkriptionshandbuch.pdf
[5] http://agd.ids-mannheim.de/html/FOLKER-Transkriptionshandbuch.pdf
[6] = Leerzeichen
[7] http://childes.psy.cmu.edu/manuals/chat.pdf
[8] = Äußerung

IPA
Das Internationale Phonetische Alphabet kann ebenso für Transkriptionen im Partitur-Editor herangezogen werden. Der IPA-Segmentierungsalgorithmus segmentiert eine Transkription, die nach IPA-Konventionen angefertigt wurde, in Wörter und Silben. Die Details dieser Konventionen wurden in:

- Thoma, Dieter & Tracy, Rosemarie (2005): L1 and Early L2: What's the difference? Talk, DGfS-Jahrestagung in Köln

veröffentlicht.

Eine schriftlich publizierte Version dieser Konventionen existiert z.Z. noch nicht. Die Konventionen sind aber, was die für die Segmentierung relevanten Zeichen anbelangt, denkbar einfach: Wörter werden mit einem Leerzeichen abgeschlossen, verschiedene Silben eines Wortes durch Punkte voneinander getrennt. Da Jim nur einsilbige Wörter verwendet, wird der Segmentierungsalgorithmus an einem anderen Beispiel verdeutlicht:



Mit der (optionalen) Segmentierung in Silben, sieht die Segmentierung folgendermaßen aus:

Segment Chain														
W			WB[9]	W			WB	W			WB	W		
SL[10]	SB[11]	SL		SL	SB	SL		SL	SB	SL		SL	SB	SL
næ:	.	ra		ɧʉ:	.	tər		ɪŋ:	.	ən		hɑ:	.	rɛ

2. Segmentierungsfehler
Wenn die Basistranskription nicht den Konventionen des Transkriptionssystems folgt, kann sie auch nicht als eine korrekte segmentierte Transkription ausgegeben werden.
Damit die segmentierte Transkription exportiert werden kann, müssen daher erst alle Fehler korrigiert sein. Um den Segmentierungsalgorithmus, den Sie unter **Bearbeiten > Voreinstellungen... > Segmentierung** festgelegt haben, anzuwenden und die Liste mit Segmentierungsfehlern zu laden, wählen Sie **Transkription > Segmentierungsfehler...**.

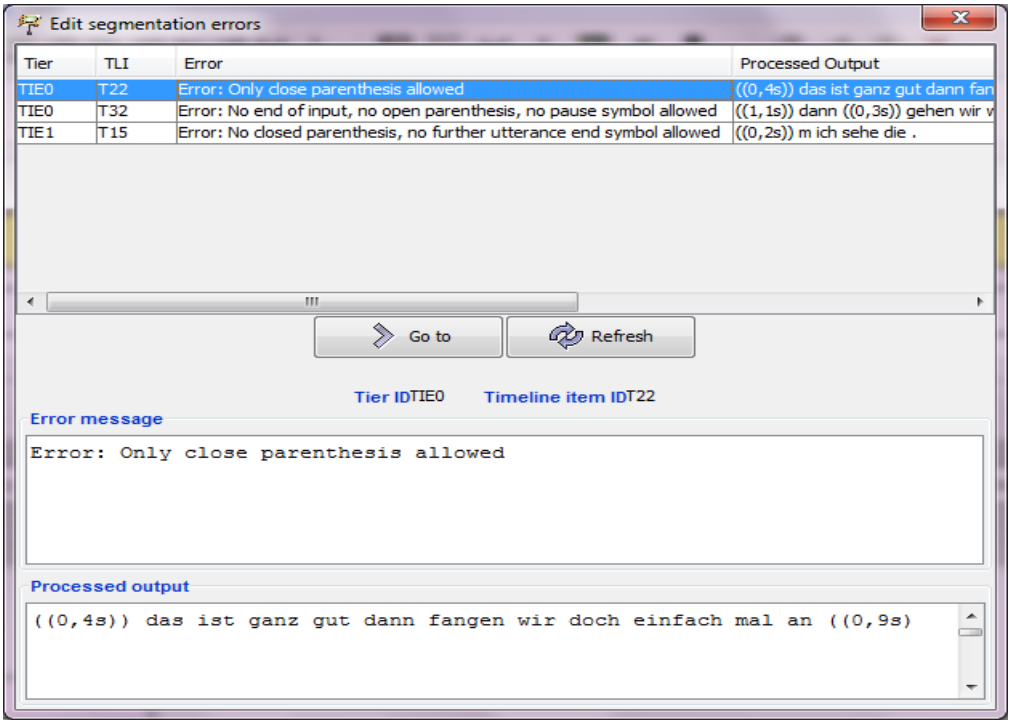
Gehen Sie sicher, dass die gewünschte Segmentierung aktiviert ist. Sollten keine Segmentierungsfehler vorliegen, können Sie den Segmentierungsalgorithmus entweder für eine Segmentzählung oder für den Export einer segmentierten Transkription nutzen.

Wenn es Segmentierungsfehler gibt, erscheint eine Liste mit folgenden Inforationen über jeden Fehler:

- Tier: Die Spur, in der der Fehler auftritt.
- TLI: Der Zeitpunkt des Fehlers in der Transkription
- Error: Eine Beschreibung des Fehlers
- Processed output: Der letzte Abschnitt, der bearbeitet und als Ausgabe umgewandelt werden konnte.

[9] = Wortgrenze
[10] = Silbe
[11] = Silbengrenze

Damit Ihnen diese Information angezeigt werden kann, klicken Sie in die Zeile. Um die Transkription von Segmentierungsfehlern zu bereinigen, gehen Sie wie folgt vor: Markieren Sie den Fehler, den Sie bearbeiten wollen, indem Sie die entsprechende Zeile der Tabelle anklicken. Klicken Sie auf **Go to**, um die Partitur an die Stelle zu bewegen, wo der Fehler aufgetreten ist. Falls nötig, konsultieren Sie vorliegendes Dokument für Informationen über Transkriptionskonventionen und Segmentierung. Beheben Sie den Fehler. Klicken Sie auf **Refresh**, um die noch verbleibenden Segmentierungsfehler anzuzeigen. Bedenken Sie Enter zu drücken, sich "aus dem Event zu bewegen" oder die Transkription zu speichern, damit die Korrekturen angewendet werden.



Generisch
Die generische Segmentierung verursacht niemals Segmentierungsfehler.

HIAT
Die gängigsten Segmentierungsfehler in HIAT werden verursacht durch:

- Den Gebrauch von drei Satzpunkten anstelle des Ellipsen Symbols. Mit **In Ereignissen ersetzen...** können Sie diese Fehler korrigieren.
- Das Setzen der schließenden Klammern eines unverständlichen Teils nach dem Äußerungsendzeichen. Die schließende Klammer muss immer vor dem Äußerungsendzeichen stehen.
- Das Vergessen einer öffnenden- oder schließenden Klammer.
- Das Vergessen des abschließenden Anführungszeichens eines Zitates.

Fehler, die nicht erkannt werden, beinhalten:

- Irrtümlicherweise in mehrere Teile gesplittete Äußerungen. Solche Äußerungen entstehen, wenn die Toplevel Segmentkette unterbrochen wird, d.h. wenn ein leere, („graues") Ereignis inmitten der Äußerung auftritt.

- Äußerungen, die mit dem Ende einer Segmentkette aufhören, anstatt ein eigenes Äußerungsendzeichen zugewiesen zu bekommen.
- An Ereignisgrenzen verbundene Wörter, die entstehen, wenn das finale Leerzeichen unterschlagen wird. Erinnern Sie sich daran, dass Sie Ereignisgrenzen innerhalb von Wörtern setzen können, ohne diese zu splitten!

DIDA
Wenn Sie mit der DIDA-Konvention arbeiten, bedenken Sie bitte, dass Großbuchstaben nur in nicht-morphemischen Äußerungen zulässig sind.

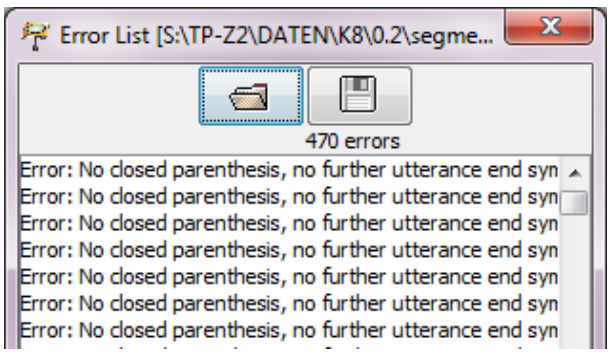
GAT und cGAT Minimal
Segmentierungsfehler entstehen in der traditionellen *GAT*-Konvention nur, wenn Endzeichen für Intonationsphrasen falsch angewendet werden. Diese sollten nicht Bestandteil von Wörtern sein.

Der Gebrauch von *cGAT* bedarf mehr Genauigkeit. Sie müssen darauf achten, dass Sie alle Konventionen korrekt anwenden. Ein Unterschied zwischen GAT und cGAT liegt in dem Gebrauch von Großschreibung und dem Gebrauch von Symbolen für Dehnungen und Betonungen.

CHAT
Auch die CHAT Konvention ist für die Segmentierung irrelevant, da Segmentierungsfehler nur aufgrund fehlerhafter Anwendung von Äußerungsendzeichen entstehen. Alle Äußerungsabschlüsse müssen gemäß den Konventionen verwendet werden.

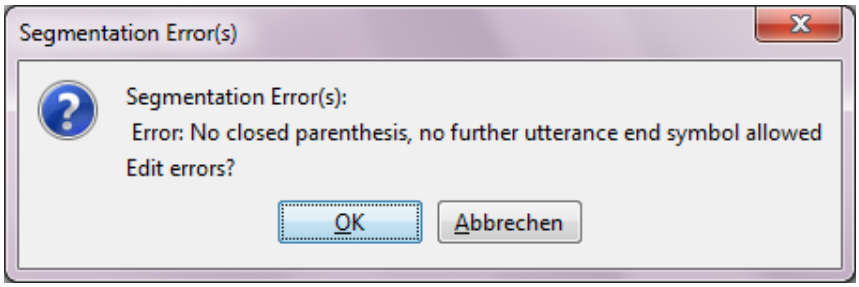
IPA
Denken Sie daran Leerzeichen zu setzen, wenn Sie Wörter voneinander trennen möchten. Punkte sollten nur für die Markierung von Silbengrenzen verwendet werden.

3. Arbeiten mit Fehlerlisten
Die Fehlerliste kann sowohl individuell im Partitur-Editor, als auch kollektiv für ein Korpus in Coma generiert werden. Genauer gesagt, werden hierbei separate Fehlerlisten für Segmentierungs- und Strukturfehler erstellt. Die Fehlerliste ist eine XML-Datei, die Informationen über die Fehler und die Transkription, in der sie auftreten, bereitstellt. Die Fehlerliste wird in den Partitur-Editor geladen, in dem man auf den Ordner in dem Dialog klickt, der sich öffnet wenn man **Datei > Fehlerliste...** anwählt. Nach dem Öffnen der Fehlerliste können Sie alle Fehler des aktuellen Korpus sehen. Mit einem Doppelklick auf den Fehler, öffnet sich die entsprechende Transkription an der Stelle, an der der Fehler auftritt. Nach der Korrektur verschwindet der Fehler aus der Liste. Wenn es zu viele Fehler sind, um sie auf einmal zu korrigieren, können Sie die Liste der übrigen Fehler jederzeit durch einen Klick auf das Disketten-Symbol abspeichern.



4. Exportieren segmentierter Transkriptionen

Um eine segmentierte Transkription mit dem spezifischen Segmentierungsalgorithmus zu erstellen, den Sie unter **Bearbeiten > Voreinstellungen** gewählt haben, klicken Sie auf **Transkription > Segmentierte Transkription exportieren...**. Sollten keine Segmentierungsfehler auftreten, speichern Sie die Datei vorzugweise unter einem Namen, der mit der Basistranskription übereinstimmt. Hierfür können sie das Suffix _s. gebrauchen. Wenn die Transkription Fehler enthält, werden Sie eine Fehlermeldung wie die folgende bekommen:



Klicken Sie **OK** um eine Liste der Segmentierungsfehler zu erhalten.

5. Segmentauszählung

Um eine Liste der Häufigkeiten der verschiedenen Segmenttypen der Transkription mit dem gewählten Segmentierungsalgorithmus zu erstellen, wählen Sie **Transkription > Transkription auszählen...**. Sollten Sie eine Fehlermeldung aufgrund von Segmentierungsfehlern erhalten, bearbeiten Sie diese wie im vorigen Abschnitt erklärt.

6. Wortlistengenerierung

Um eine Wortliste zu generieren, wählen Sie **Wortliste...**im Menüpunkt **Transkription**. Die Wortliste wird in einem neuen Fenster angezeigt und beinhaltet alle Einheiten als segmentierte Wörter in der Transkription. Durch Anklicken von **Sprecher** und **Wörter** können Sie die Liste je nach Auswahl alphabetisch sortieren lassen Um die Wortliste als HTML-Datei zu speichern, gehen Sie auf Speichern unter…. Sie haben folgende Optionen: Wählen Sie **Simple word list** (HTML) für eine Liste mit den alphabetisch geordneten Wörtern oder **Word list by speaker** (HTML) für eine Liste, in der die Worten anhand der Sprechern geordnet werden. Sollten Sie eine Meldung über Segmentierungsfehler bekommen, bearbeiten Sie diese wie im vorigen Abschnitt erklärt.

Appendix: GAT2 Transkriptionskonventionen für Minimaltranskripte

Thomas Schmidt

N.B.: *GAT2 ist eine Transkriptionskonvention, die für das Deutsche entwickelt wurde. Einige der Konventionen lassen sich nicht einfach ins Englische übertragen, da Sie aufgrund der deutschen Orthographie entwickelt wurden. Dieses Dokument gibt einen praktischen Überblick über die Transkriptionskontrolle von GAT2 fürs Englische in Folker und stellt keine „richtige" Transkriptionskonvention dar. Wenn Sie anderen Konventionen folgen möchten, wählen Sie* Bearbeiten > Voreinstellungen > Transkriptionsstufe *an, um die Kontrolle auszuschalten.*

1. Wörter

- Keine Großschreibung **wayne rooney**, nicht: Wayne Rooney
- Zahlen werden ausgeschrieben. Abkürzungen werden nicht gebraucht. **elf**, nicht: 11
  Doktor, nicht: Dr.
- Keine Zeichensetzung (d.h.: keine Bindestriche in Wörtern, keine Satzpunkte in Abkürzungen, keine Apostrophe, wählen Sie für Kontraktionen stattdessen einen Unterstrich):
  they_re, nicht: they're

2. Pausen

- Pausen (ausgenommen Mikropausen) werden vorzugweise in separaten Segmenten transkribiert
- Pausen (ausgenommen Mikropausen) sind vorzugweise keinem Sprecher zugeordnet
- Gemessene Pausen werden als Zahl mit Dezimalpunkt und zwei Nachkommastellen in Klammern gesetzt.
  (0.85)
- Mikropausen (Pausen, die zu kurz für eine Messung sind) werden geschrieben als
  (.)

3. Unverständliches/ schwer Verständliches

- Vermutungen (werden durch einfache Klammern angezeigt. Es dürfen nur ganze Wörter oder Wortketten in Klammern gesetzt werden. **(whatever), not what(ever)**
- Alternativvorschläge können, durch Schrägstrich getrennt, in die Klammern eingefügt werden.
  (cat/bat/mat)
- Unverständliche Passagen werden als +++ transkribiert. Die Anzahl der +++ entspricht der Anzahl der (vermuteten) Silben.
  ++++++ +++
- Längere unverständliche Passagen oder Passagen, in denen die Silben nicht ausgemacht werden können, werden transkribiert als
  ((unverständlich))

4. Gemischtes

- Hörbares Atmen wird mit dem Grad-Symbol transkribiert (beim Einatmen vorgestellt, beim Ausatmen nachgestellt) und ein bis drei Einheiten des Buchstabens H entsprechend der Länge des Ereignisses.
  °hhh
  hh°
- Kontraktionen werden durch einen Unterstrich angezeigt
  we_re
- Nicht phonologische Segmente werden in doppelte Klammern gesetzt.
  ((lacht))