

## POSTPRINT

Thomas Schmidt

### 27. Werkzeuge für die Transkription gesprochener Sprache

Transkriptionswerkzeuge sind spezialisierte Softwaretools für die Transkription und Annotation von Audio- oder Videoaufzeichnungen gesprochener Sprache. Dieses Kapitel erklärt einleitend, worin der zusätzliche Nutzen solcher Werkzeuge gegenüber einfacher Textverarbeitungssoftware liegt, und gibt dann einen Überblick über grundlegende Prinzipien und einige weitverbreitete Tools dieser Art. Am Beispiel der Editoren FOLKER und OrthoNormal wird schließlich der praktische Einsatz zweier Werkzeuge in den Arbeitsabläufen eines Korpusprojekts illustriert.

#### I. Einleitung: Warum spezialisierte Werkzeuge?

Für die Transkription von gesprochener Sprache, die als digitales Audio oder Video aufgezeichnet vorliegt, wird spezialisierte Software benutzt, über die dieses Kapitel erstens einen Überblick geben möchte und deren Verwendungsweise hier zweitens anhand eines konkreten Beispiels illustriert werden soll.

Transkription (→ Kapitel 23 [Gesprächsanalytische Transkription] in diesem Band) war ein wichtiger Bestandteil linguistischer Methodik, lange bevor Computer oder gar spezialisierte digitale Werkzeuge für diese Aufgabe zur Verfügung standen. Beispielsweise wurde das zwischen 1955 und 1970 entstandene Korpus „Deutsche Mundarten“ (auch bekannt als „Zwirner-Korpus“, → Kapitel 25 [Korpora gesprochener Sprache] in diesem Band) ursprünglich hand- oder maschinenschriftlich transkribiert, die Auswertung der Daten erfolgte dann anhand der so entstandenen Papiertranskripte. Und auch im digitalen Zeitalter lässt sich die Aufgabe der Transkription prinzipiell auch ohne spezialisierte Software, beispielsweise mit einem Texteditor oder Textverarbeitungsprogramm wie Word, angehen.

Was sind also Gründe, spezialisierte Transkriptionssoftware zu benutzen? Gesprochene Sprache hat *erstens* eine Reihe von Eigenschaften, die ihre Repräsentation komplexer macht als das, was wir üblicherweise in der Struktur eines schriftlichen Textes abbilden

können. Gesprochene Sprache ist nicht rein linear, sondern enthält parallele Strukturen (simultanes Sprechen, „Überlappung“, Gleichzeitigkeit von verbalem und non-verbalem Handeln); die Zuordnung des Gesprochenen zu den am Gespräch beteiligten Sprecher\*innen ist wesentlich für ein Verständnis des Gesprächsablaufs; neben geäußerten Wörtern enthält mündliche Interaktion auch Stille (Pausen) oder Nonverbales (z. B. Lachen, Nicken), das zum Verständnis und zur Analyse wichtig sein kann und im Transkript von transkribierten Wortformen unterschieden werden muss. Ein Transkript ist *zweitens* kein autonomer Text, sondern über zeitliche Verweise mit der zugrunde liegenden Audio- oder Videoaufnahme verbunden („aligniert“). *Drittens* ist die reine Transkription des Gesagten oft nicht die alleinige Analysegrundlage eines sprachwissenschaftlichen Forschungsprojekts, sondern es werden ausgewählte transkribierte „Text“-Bestandteile mit zusätzlichen analytischen Informationen (Annotationen wie z. B. Part-of-Speech-Tagging [POS-Tagging], → Kapitel 14 [Annotation] in diesem Band) angereichert. Spezialisierte Transkriptionssoftware berücksichtigt diese strukturelle Komplexität: Sie repräsentiert das Transkript nicht als einfachen linearen Text, sondern in einer Datenstruktur, in der Simultaneität, Sprecherzuordnung, die Unterscheidung von verbalem und non-verbalem Material, die Verknüpfung zur zugrunde liegenden Aufnahme und zusätzliche Annotationen so abgebildet wer-

den können, dass sie im Forschungsprozess optimal genutzt werden können.

Die so entstehenden Datenstrukturen können also viel besser computergestützt weiterverarbeitet werden, als das bei einer Textdatei möglich wäre, die nichts von der spezifischen Struktur des Transkripts, das sie repräsentiert, „weiß“. Wichtige solcher Verarbeitungsmöglichkeiten sind:

- die Arbeit mit Transkripten unter Ausnutzung des Text-Ton-Alignments, also des direkten Rückgriffs von einer Stelle im Transkript auf die entsprechende Stelle in der Audio- oder Videoaufnahme;
- die flexible Ausgabe von Transkripten in verschiedenen Darstellungsweisen und Formaten – z. B. als Partitur oder Liste, als HTML-Datei zur Anzeige in einem Browser oder als Textdatei zur Integration in eine Publikation;
- der Export von Transkripten in Formate zur Weiterbearbeitung in anderer Software, z. B. zur phonetischen Analyse in Praat (s.u.) oder zur statistischen Auswertung mit R, SPSS oder ähnlichen Tools (→ Kapitel 28 [Werkzeuge statistische Analyse] in diesem Band);
- die automatische Annotation von Transkripten mit geeigneten computerlinguistischen Tools wie z. B. einem POS-Tagger (→ Kapitel 30 [Werkzeuge automatische Sprachanalyse] in diesem Band);
- die systematische Recherche (*Query*) in Transkripten und Annotationen in einer korpuslinguistischen Herangehensweise.

Zudem kann spezialisierte Software auf einer ganz praktischen Ebene den aufwändigen Transkriptionsprozess effektiver bzw. effizienter machen. Anfänglich bedeutet das Erlernen einer spezialisierten Software zwar zusätzlichen Aufwand. Wenn die Handgriffe zur Bedienung solcher Software erst einmal eingeübt sind, lässt sich damit aber oft schneller, in jedem Falle aber präziser transkribieren als mit alternativen Lösungen.

## 2. Grundlagen

Transkriptions-Werkzeuge arbeiten alle nach ähnlichen Prinzipien, weisen aber unterschiedliche Spezialisierungen auf. Die hier vorgestellten Tools sind untereinander weitestgehend interoperabel, d. h., es können Daten relativ leicht zwischen ihnen ausgetauscht werden.

### 2.1 Partiturdarstellung

Das allen Transkriptionswerkzeugen gemeinsame Prinzip kann wie folgt beschrieben werden: Die Transkription erfolgt, indem wiederholt ein Ausschnitt aus einer Audio- oder Videodatei ausgewählt, diesem eine Zeichenkette (die eigentliche Transkription) zugewiesen und in eine vorgegebene Struktur eingeordnet wird. So entsteht ein Transkriptions-„Text“, der stückweise mit der zu Grunde liegenden Audio- oder Videodatei aligniert ist.

Die gängigste und flexibelste Struktur für die Darstellung von Transkripten in Transkriptionstools ist die einer Partitur. Wie in einer musikalischen Partitur, in der verschiedene untereinanderstehende Notensysteme das Zusammenspiel verschiedener Instrumente abbilden, werden in einer Transkriptions-Partitur die Handlungen verschiedener Sprecher\*innen in der Interaktion in ihrer Abfolge und Gleichzeitigkeit in verschiedenen „Spuren“ dargestellt. So lassen sich z. B. simultanes Sprechen oder das gleichzeitige Auftreten von verbalen und non-verbalen Handlungen intuitiv darstellen. Nach dem gleichen Prinzip können dem transkribierten Material durch weitere Spuren Informationen in Form von Annotationen hinzugefügt werden. Das Beispiel in Abbildung 1 illustriert dies.

Hier sind verbale und non-verbale Handlungen zweier Sprecher\*innen – bezeichnet mit den Kürzeln DS und FB – abgebildet. Von links nach rechts gelesen ergibt sich der zeitliche Ablauf der Interaktion: DS beginnt, FB fällt ihr / ihm ins Wort. Von beiden Sprecher\*innen simultan geäußerte Redebestandteile („d'accord“ und „Alors ça“) stehen horizontal an der gleichen Position untereinander. In gleicher Weise wird die zeitliche Ausprägung

DS [sup]	faster
DS [l]	Okay, D'accord, d'accord
DS [en]	Okay, Agreed, agreed
DS [st]	right hand raised
FB [l]	Alors ça dépend ((cough)) un petit peu.
FB [en]	That depends then, a little bit.
DS [pho]	stpe

Abb. 1: Partitur-Darstellung eines Transkripts mit verschiedenen Annotationsspuren

einer non-verbalen Handlung von Sprecher\*in DS („right hand raised“) abgebildet. Annotationen auf verschiedenen Spuren beschreiben suprasegmentale Eigenschaften des Gesagten (Sprechgeschwindigkeit „faster“ in Spur DS [sup]), liefern Übersetzungen des Transkribierten vom Französischen ins Englische (Spuren DS [en] und FB [en]) oder präzisieren Aussprachebesonderheiten in phonetischer Umschrift (Spur FB [pho]).

## 2.2 Interoperabilität

Eine Eigenschaft, die bei der Auswahl eines Transkriptionswerkzeugs eine wichtige Rolle spielen sollte, ist dessen Interoperabilität, also die Möglichkeit, die mit dem Werkzeug erstellten Daten auch in anderen Werkzeugen nutzen zu können. Interoperabilität ist vor allem dann nützlich, wenn ein einzelnes Werkzeug nicht alle erforderlichen Funktionen zur Verfügung stellt. Beispielsweise kann ELAN (EUDICO Linguistic Annotator, s. Abschnitt 3.2) benötigt werden, um Videos zu transkribieren, eine detaillierte Annotation prosodischer Aspekte aber die betreffenden Funktionen in Praat (s. Abschnitt 3.3) erfordern. In einem geeigneten Workflow kann die Interoperabilität der beiden Werkzeuge genutzt werden, um ihre jeweiligen Stärken miteinander zu kombinieren.

Da es noch keinen allgemein unterstützten Standard für die Datenformate von Transkripten gibt, erfolgt der Datenaustausch üblicherweise über direkte Datenexporte und -importe der jeweiligen Tools. Wie die Abbildungen 2 bis 5, die jeweils den gleichen Transkriptausschnitt in den unterschiedlichen Werkzeugen zeigen, bereits andeuten,

ist eine weitreichende Interoperabilität zwischen EXMARaLDA/FOLKER, ELAN/Simple ELAN und Praat (und bis zu einem gewissen Grad auch CLAN [Computerized Language Analysis]) gegeben. Da diese Tools mit sehr ähnlichen Datenmodellen arbeiten, ist ein Datenaustausch zwischen ihnen mit nur wenig Informationsverlust möglich. Mögliche Informationsverluste beim Wechsel zwischen den Werkzeugen sollte man bei der Entwicklung des Korpus-Workflows berücksichtigen.

Über den Datenaustausch zwischen den Transkriptionswerkzeugen hinaus gibt es in den einzelnen Tools auch Exportmöglichkeiten, die eine Weiterverarbeitung transkribierter Daten mit anderer Software ermöglichen. Dazu gehören insbesondere Exporte in Text- oder Textverarbeitungsformate, mit deren Hilfe z. B. Transkriptausschnitte in eine Publikation integriert werden können.

## 3. Kurzdarstellung ausgewählter Werkzeuge

In den Sprachwissenschaften hat eine überschaubare Anzahl von Transkriptions- und Annotations-Werkzeugen, die im Folgenden vorgestellt werden, weite Verbreitung gefunden.

### 3.1 EXMARaLDA

Das EXMARaLDA-System (Schmidt/Wörner 2014) wurde ursprünglich für die Arbeiten am Sonderforschungsbereich „Mehrsprachigkeit“ entwickelt. Es wird vor allem in der Gesprächs- und Variationsforschung genutzt. EXMARaLDA besteht aus mehreren Werkzeugen, die über die Transkription und Annotation hinaus auch die Verwaltung und die Analyse von Korpora gesprochener Sprache unterstützen. Das am breitesten einsetzbare Werkzeug ist der EXMARaLDA Partitur-Editor, der die Transkription und Annotation von Audio- oder Videodaten in Partitur-Darstellung ermöglicht. Abbildung 2 zeigt das Benutzerinterface des Partitur-Editors: im oberen Bereich wird die Audio-Datei als Os-

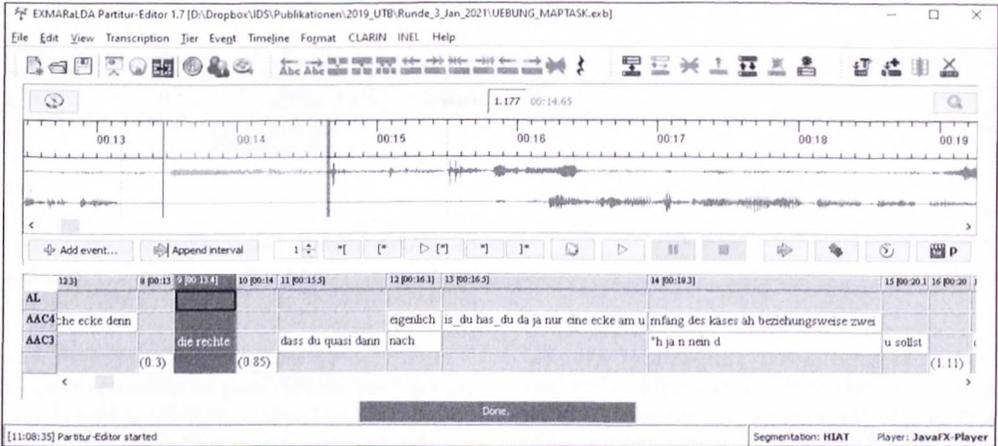


Abb. 2: Benutzerinterface des EXMARaLDA Partitur-Editors

zillogramm visualisiert, in dem Abschnitte der Aufnahme ausgewählt werden können. Der untere Bereich stellt das Transkript als Partitur dar.

Neben dem Partitur-Editor beinhaltet das EXMARaLDA-System weitere Werkzeuge: Mit FOLKER (FOLK-Editor) wurde auf der Grundlage des Partitur-Editors ein zusätzlicher Transkriptions-Editor entwickelt, der für die Arbeitsabläufe im FOLK-Projekt<sup>1</sup>, also für eine effiziente Transkription nach den Konventionen für cGAT-Minimaltranskripte, optimiert ist (vgl. Schmidt/Schütte 2010, und s.u.). Ebenfalls für das FOLK-Projekt entwickelt wurde OrthoNormal, ein Tool, mit dem transkribierte Daten mit orthographisch normalisierten Formen annotiert werden können (s. Abschnitt 4.2). Neben diesen Transkriptions- und Annotationswerkzeugen beinhaltet EXMARaLDA mit dem Corpus-Manager (Coma) ein Werkzeug zum Verwalten von Korpora gesprochener Sprache und mit dem EXMARaLDA Analyse- und Konkordanz-tool (EXAKT) ein Werkzeug für die Suche auf und Analyse von solchen Korpora.<sup>2</sup>

### 3.2 ELAN

ELAN (Wittenburg et al. 2006) wird am Max-Planck-Institut für Psycholinguistik in Nijmegen entwickelt. Zwei der wichtigsten Bereiche, in denen das Werkzeug breitere Verwendung gefunden hat, sind die Dokumentation bedrohter Sprachen (dies ursprünglich vor allem im Kontext des Projekts DOBES<sup>3</sup>) und die Gebärdensprachforschung. Da in diesen Bereichen typischer- oder gar notwendigerweise mit Video-Daten gearbeitet wird, ist die Video-Funktionalität von ELAN besonders weit ausgebaut, beispielsweise können hier zwei oder mehr Videos gleichzeitig abgespielt werden. Wie in Abbildung 3 illustriert, besteht das Benutzerinterface von ELAN aus den gleichen Komponenten – Oszillogramm für die Visualisierung der Audio-Datei, Partitur für die Darstellung des Transkripts – wie der EXMARaLDA-Partitur-Editor.

So wie FOLKER als Variante des EXMARaLDA-Partitur-Editors entwickelt wurde, hat auch ELAN mit dem Werkzeug Simple-ELAN eine Variante mit reduziertem Funktionsumfang, die für eine einfache und effiziente Transkription optimiert ist.<sup>4</sup>

1 <http://agd.ids-mannheim.de/folk.shtml>.

2 Alle EXMARaLDA-Werkzeuge können kostenlos von <http://www.exmaralda.org> heruntergeladen werden.

3 Dokumentation bedrohter Sprachen, <https://dobes.mpi.nl/>.

4 Beide Tools können von der Webseite des Language Archive in Nijmegen kostenlos heruntergeladen werden: <https://tla.mpi.nl/tools/tla-tools/elan/>.

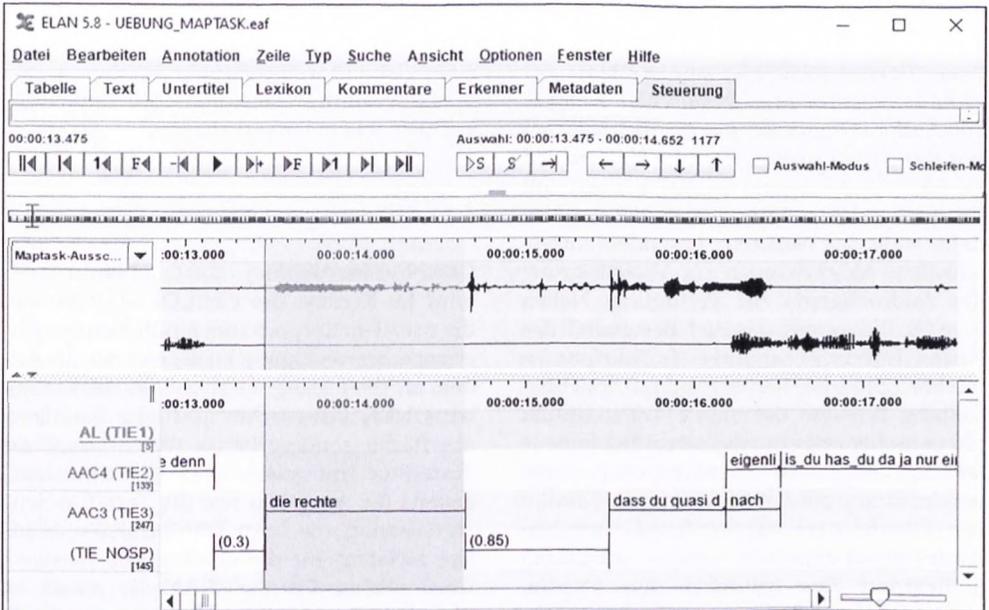


Abb. 3: Benutzerinterface von ELAN

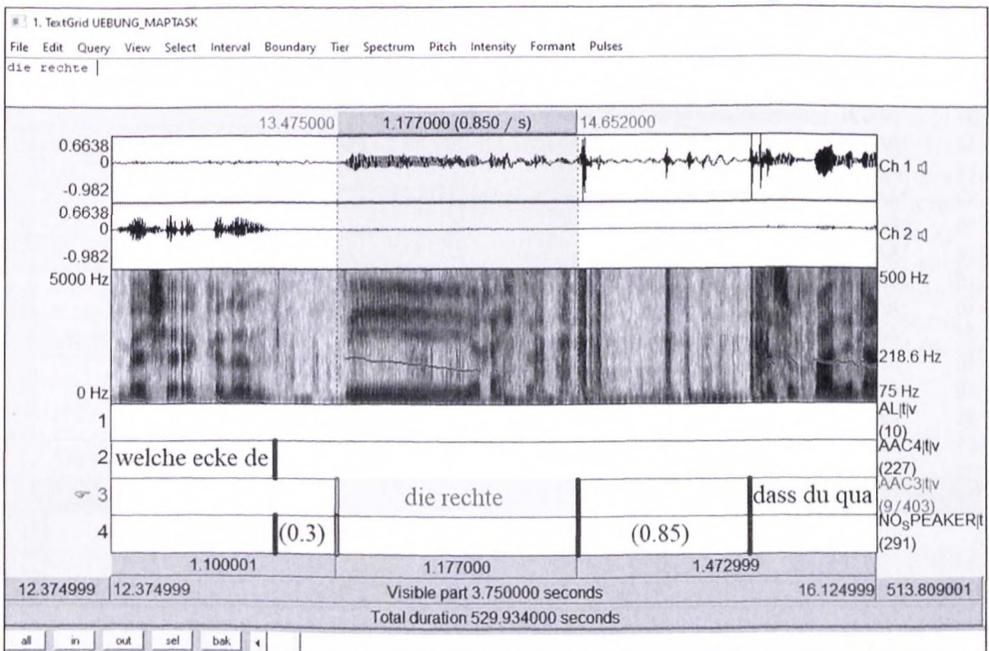


Abb. 4: Benutzerinterface von Praat

### 3.3 Praat

Praat (Boersma 2001) ist eine Software zur computergestützten phonetischen Analyse von Audio-Daten, wird aber oft auch als Transkriptionswerkzeug genutzt. Im Unterschied zu EXMARaLDA und ELAN können mit Praat keine Video-Daten bearbeitet werden, dafür stellt das Werkzeug besonders fortschrittliche Möglichkeiten zur Visualisierung des Audio-Signals zur Verfügung. Neben dem Oszillogramm, das auch Bestandteil des Nutzer-Interfaces anderer Transkriptionswerkzeuge ist, können in Praat z. B. als Hilfestellung für eine detaillierte Transkription phonetischer oder prosodischer Phänomene auch Spektrogramm-Darstellungen, Formanten-Analysen oder Pitch-Konturen (Tonhöhenverläufe) angezeigt werden, die aus dem Audio-Signal extrahiert werden. Abbildung 4 illustriert dies: Zwischen dem Oszillogramm im oberen Bereich und der Partitur im

unteren Bereich enthält das Benutzerinterface von Praat eine dritte Komponente, die hier eine Visualisierung des Audio-Signals als Spektrogramm (Graustufen) und Tonhöhenverläufe (blaue Linien) anzeigt.<sup>5</sup>

### 3.4 CLAN

CLAN (MacWhinney 2000, Abbildung 5) wird im Kontext der CHILDES-Datenbank, die vor allem Korpora zum kindlichen Spracherwerb zur Verfügung stellt, entwickelt. Anders als die bislang vorgestellten Werkzeuge hat CLAN keine echte grafische Benutzeroberfläche, sondern ist im Wesentlichen ein Texteditor mit zusätzlicher Funktionalität, erstens für Aufgaben wie die Text-Ton-Synchronisation, die beim Transkribieren anfallen, zweitens für die umfangliche Auswertung solcher Daten. CLAN ist damit in vielerlei Hinsicht ein eher veraltetes Tran-

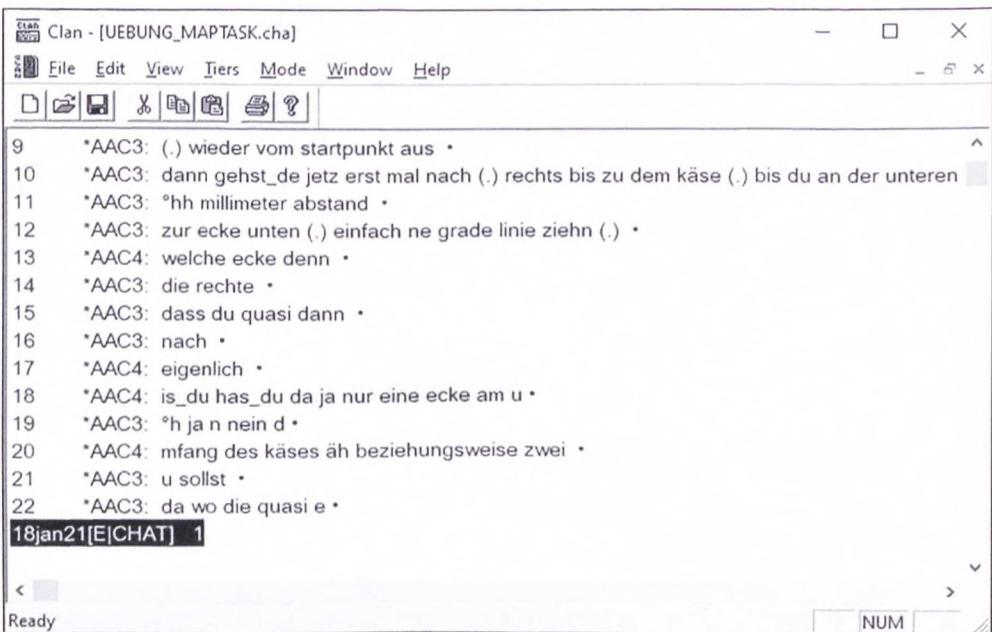


Abb. 5: Benutzerinterface von CLAN

5 Praat kann von der Website <http://www.fon.hum.uva.nl/praat/> kostenlos heruntergeladen werden.

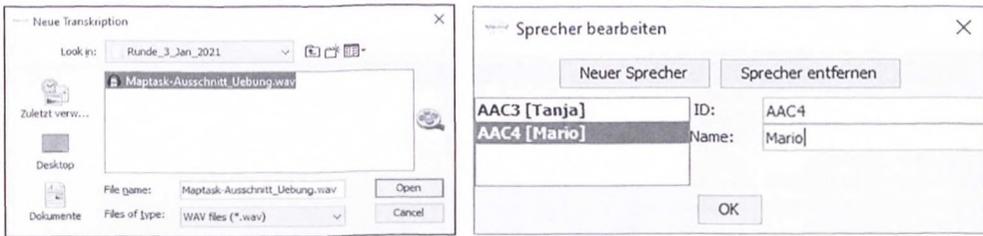


Abb. 6: Auswahl einer Audio-Datei und Anlegen der Sprechertabelle

skriptionswerkzeug. Es wird dennoch nach wie vor häufig verwendet, besonders in seinem ursprünglichen Anwendungsbereich, der Spracherwerbsforschung.<sup>6</sup>

## 4. Arbeitsschritte

In diesem Abschnitt zeigen wir, wie Transkriptions- und Annotationswerkzeuge im Workflow des FOLK-Projektes eingesetzt werden, um ausgehend von einer Audio- oder Videoaufnahme schrittweise ein publikationsfähiges Transkript zu erstellen, das

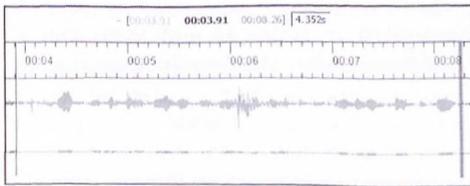


Abb. 7: Auswahl eines Aufnahmeabschnitts im Oszillogramm

neben der eigentlichen Transkription zusätzliche Annotationen auf mehreren Ebenen beinhaltet. Wie oben bereits erwähnt, arbeitet das FOLK-Projekt mit zwei Werkzeugen, die Teil des EXMARaLDA-Systems sind: dem Transkriptionseditor FOLKER und dem Annotationswerkzeug OrthoNormal.

### 4.1 Transkription

Zum Anlegen eines neuen Transkripts muss in FOLKER zunächst die zu transkribierende Audio-Datei ausgewählt werden (s. Abb. 6). Anschließend werden in einer Sprechertabelle die am Gespräch beteiligten Sprecher\*innen definiert.

Die ausgewählte Audio-Datei wird in FOLKER dann als Oszillogramm angezeigt. Darin wird ein geeigneter Ausschnitt – typischerweise von einer Länge von bis zu fünf Sekunden – zum Transkribieren ausgewählt (s. Abb. 7). In der Segmentansicht wird für den ausgewählten Abschnitt dann ein neuer Eintrag mit dem betreffenden Start- und Endpunkt angelegt. In einem Textfeld kann zunächst die ei-

	Start	Ende	Sprecher	Transkriptionstext	Syntax	Zeit
1	00:00.48	00:02.22	---	wieder vom startpunkt aus	✓	✓
			AAC3			
			AAC4			
			---			
			Tanja			

Abb. 8: Eingabe von Transkripttext und Sprecherzuweisung

6 CLAN kann kostenlos von <http://dali.talkbank.org/clan/> heruntergeladen werden.

	Start	Ende	Sprecher	Transkriptionstext	Syntax	Zeit
1	00:00.0		AAC3	also		
2	00:00.45	00:01.57		(.) wieder vom startpunkt aus	✓	✓
3	00:01.87	00:02.21		(0.34)		
4	00:02.21	00:08.25	AAC3	dann gehst_de jetzt erst mal nach (.) rechts bis zu dem kase (.) bis du an der unteren ecke vom kas...		
5	00:08.25	00:09.80	AAC3	°hh millimeter abstand		
6	00:09.80	00:10.05		(0.25)		
7	00:10.05	00:12.37	AAC3	zur ecke unten (.) einfach ne grade linie ziehn (.)		
8	00:12.37	00:13.17	AAC4	welche ecke denn		
9	00:13.17	00:13.47		(0.3)		
10	00:13.47	00:14.65	AAC3	die rechte		
11	00:14.65	00:15.50		(0.85)		
12	00:15.50	00:16.19	AAC3	dass du quasi dann		
13	00:16.19	00:16.51	AAC4	eigentlich		
14	00:16.19	00:16.51	AAC3	nach		
15	00:16.51	00:18.34	AAC4	is_du_has_du da ja nur eine ecke am u		
16	00:18.34	00:20.17	AAC4	mfang des kases ah beziehungsweise zwei		

Abb. 9: Vollständiges Transkript in der FOLKER Segment-Ansicht

Segmente	Partitur	Beiträge
		7 [00:12.3] 8 [00:13] 9 [00:13.4] 10 [00:14] 11 [00:15.5] 12 [00:16.3] 13 [00:16.5] 14 [00:18.3]
AAC3	zieh (.)	die rechte
AAC4	welche ecke denn	dass du quasi dann
AL		eigentlich
		is_du_has_du da ja nur eine ecke am u
		mfang des kases ah beziehungsweise zw
		(0.3) (0.85)

Abb. 10: Partitur-Darstellung des Transkripts in FOLKER

gentliche Transkription eingegeben werden. In der zusätzlichen Spalte „Syntax“ wird geprüft, ob der eingegebene Text den Vorgaben der cGAT-Konventionen für Minimaltranskripte folgt. Bei erfolgreicher Prüfung erscheint dort ein grünes Häkchen, bei einem korrekturbedürftigen Verstoß gegen die Konventionen (z.B. bei der Verwendung eines Großbuchstabens) ein rotes Kreuz.

Anschließend wird dieses Segment einem der zuvor definierten Sprecher\*innen zugewiesen (s. Abb. 8).

Indem diese Handgriffe- Auswahl eines Ausschnitts in der Aufnahme, Anlegen des zugehörigen Segments, Eingabe des Transkriptionstexts und Sprecherzuweisung – nun vielfach wiederholt werden, entsteht nach und nach ein vollständiges Transkript (s. Abb. 9).

Zeitliche Parallelität kann dabei in der Segmentansicht von FOLKER zunächst nur an den Start- und Endzeiten der einzelnen Seg-

mente (z.B. in Zeilen 13 und 14 in Abb. 8) abgelesen werden. Um Simultanpassagen direkter visuell erfassen und präzise bearbeiten zu können, kann in FOLKER in eine Ansicht gewechselt werden, in der das gleiche Transkript als Partitur angezeigt wird (s. Abb. 10).

Die Ersttranskription in FOLK ist abgeschlossen, wenn für die Aufnahme eine lückenlose Liste transkribierter Segmente erstellt und der Transkriptionstext vollständig auf Konformität mit den cGAT-Konventionen geprüft wurde. In der Beitrags-Ansicht von FOLKER erfolgt dann eine abschließende inhaltliche Überprüfung und ggf. Korrektur des Transkripts.

#### 4.2 Orthographische Normalisierung

Für die eigentliche Transkription in FOLK wird gemäß den cGAT-Konventionen die sogenannte literarische Umschrift verwendet,

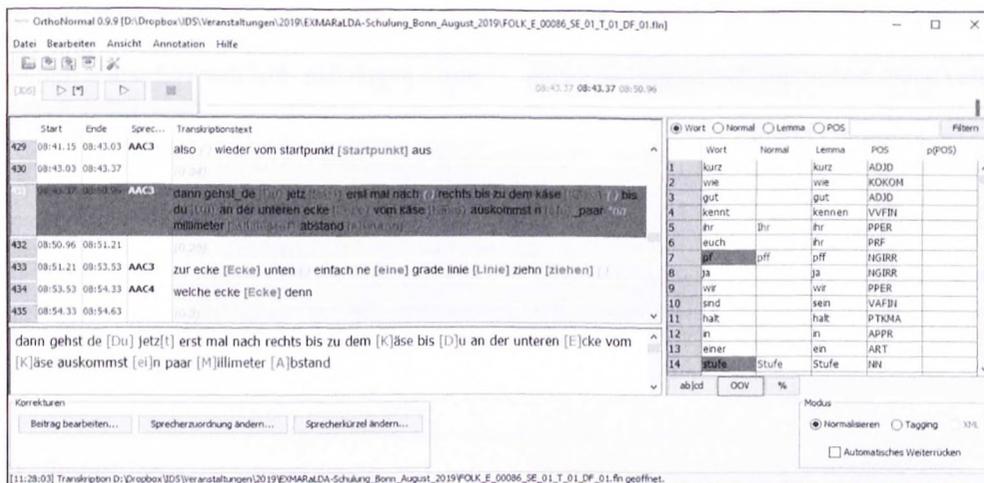


Abb. 11: Benutzerinterface von OrthoNormal

die abweichende Lautungen, seien sie allgemeinen Mechanismen der Mündlichkeit wie Elisionen („ich hab“ statt „ich habe“), Verschleifungen („haste“ statt „hast Du“), Versprechern oder dialektaler Variation („haschd“ statt „hast“) geschuldet, berücksichtigt und nach den Laut-Graphem-Regelmäßigkeiten der Orthographie möglichst gut abzubilden versucht. Dies vermittelt beim Lesen des Transkripts ein besseres Abbild der tatsächlich geäußerten Formen. Für andere Einsatzbereiche des Transkripts, wie einer Korpus-Suche oder der automatischen Anreicherung durch weitere Annotationen, ist es aber nachteilig, weil sich die Formenvielfalt so deutlich erhöht und damit weniger gut kontrollieren lässt. In FOLK wird die ursprüngliche literarische Transkription daher auf einer zweiten Annotationsebene um eine sogenannte orthographische Normalisierung erweitert, die jeder davon abweichenden Form die zugehörige Form aus der Standardorthographie zuordnet. Für die effiziente orthographische Normalisierung wurde mit OrthoNormal ein eigener Editor entwickelt. Mit dessen Hilfe wird ein abgeschlossenes Transkript zunächst automatisch normalisiert, indem bereits bekannte Normalisierungs-paare aus einem Wörterbuch abgerufen werden. Dieser Prozess ist mit einer Fehlerquote von etwa 20% behaftet (vgl.



Abb. 12: Normalisierungsdialog für das transkribierte Token ‚de‘

Schmidt 2016) und wird daher um eine manuelle Normalisierung des Transkripts ergänzt. Hierfür wird das Transkript in einer Beitragsansicht repräsentiert, zusätzlich listet eine Tabelle alle transkribierten Wortformen mit ihren normalisierten Entsprechungen auf (s. Abb. 11).

Der manuelle Annotationsprozess besteht nun darin, für jede transkribierte Wortform zu prüfen, ob eine orthographische Normalisierung notwendig ist und diese ggf. über einen Dialog einzugeben. Das Tool stellt dabei bereits bekannte Normalisierungsformen als Auswahl zur Verfügung (s. Abb. 12).

Ein auf diese Weise vollständig normalisiertes Transkript kann mit OrthoNormal

dann automatisch mit zwei zusätzlichen Annotationsebenen angereichert werden, indem über einen Aufruf des TreeTaggers zu jedem normalisierten Token ein Lemma und ein POS-Tag ermittelt wird.

## 5. Praktisches

Alle hier vorgestellten Transkriptionstools können von der jeweiligen Homepage kostenlos heruntergeladen werden. Es werden

jeweils Versionen für unterschiedliche Betriebssysteme – Windows, MacOS und Linux – angeboten. Auf den Webseiten finden sich auch verschiedene Hilfestellungen, die in den Umgang mit dem Werkzeug einführen (z. B. Schnellstart-Anleitungen, Video-Tutorials) und spezifische Fragen zu deren Verwendung beantworten (Benutzerhandbücher, „How-To“-Dokumente). Es empfiehlt sich, sich mit diesen Hilfestellungen vertraut zu machen, um sicherzugehen, dass das Werkzeug effizient und korrekt angewendet wird.

### Zum Weiterlesen

Einen nicht mehr ganz aktuellen, aber immer noch hilfreichen Überblick über verschiedene Transkriptions- und Annotationswerkzeuge geben Rohlfig et al. (2006).

In Schmidt (2016) wird der vollständige Workflow des FOLK-Projekts in seinen technischen Details beschrieben.

Im Oxford Handbook of Corpus Phonology (Durand et al. 2014) finden sich ausführlichere Darstellungen einiger der hier vorgestellten sowie weiterer Transkriptionswerkzeuge.

## Literatur

- Boersma, Paul (2001): Praat, a system for doing phonetics by computer, in: *Glott International*, Jg. 5, H. 9/10, S. 341–345.
- Durand, Jacques, Ulrike Gut, Gjeert Kristoffersen (eds.) (2014): *The Oxford Handbook of Corpus Phonology*, Oxford: OUP.
- MacWhinney, Brian (2000): *The CHILDES project. Tools for analyzing talk*, 3. Aufl., Mahwah, NJ: Lawrence Erlbaum Associates.
- Rohlfig, Katharina, Daniel Loehr, Susan Duncan, Amanda Brown, Amy Franklin, Irene Kimbara, Jan-Torsten Milde, Fey Parrill, Travis Rose, Thomas Schmidt, Han Sloetjes, Alexandra Thies und Sandra Wellinghoff (2006): Comparison of multimodal annotation tools, in: *Gesprächsforschung*, Jg. 7, S. 99–123.
- Schmidt, Thomas (2016): Construction and Dissemination of a Corpus of Spoken Interaction – Tools and Workflows in the FOLK project, in: *Journal for Language Technology and Computational Linguistics*, Bd. 31, H. 1, S. 127–154, [online] <https://jcl.org/content/2-allissues/5-Heft1-2016/jcl-2016-1-7Schmidt.pdf>.
- Schmidt, Thomas und Wilfried Schütte (2010): FOLKER. An Annotation Tool for Efficient Transcription of Natural, Multi-party Interaction, in: Calzolari, Nicoletta et al. (eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 10)*, Valletta, S. 2091–2096, [online] <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/2232>.
- Schmidt, Thomas und Kai Wörner (2014): EXMARALDA, in: Durand, Jacques, Ulrike Gut und Gjeert Kristoffersen (eds.), *The Oxford Handbook of Corpus Phonology*, Oxford: OUP, S. 402–419.
- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann und Han Sloetjes (2006): ELAN. A Professional Framework for Multimodality Research, in: Calzolari, Nicoletta et al. (eds.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Genua, S. 1556–1559, [online] [https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item\\_60436](https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_60436).

Die Adressen aller Webseiten und Online-Ressourcen in diesem Beitrag wurden zuletzt auf Aktualität überprüft am 9. Juni 2021.