

# WEBBASIERTE TOOLS FÜR DIE TRANSKRIPTION UND ANALYSE VON REDEN

## Hilfreiche Instrumentarien für die (Polito)Linguistik

MARCELLA PALLADINO  
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

**Abstract** – To conduct a linguistic analysis of political speeches, it is essential to obtain their transcription. However, not all speeches delivered by political representatives are readily transcribed and accessible. Moreover, transcribing speeches that last more than some minutes is time-consuming and for this reason written texts, short political speeches, or speeches that already have an official transcription are more likely to be considered by linguistic disciplines. Tools and technology developed by the speech processing community facilitate the transcription process. The audio signal tends to be neglected in some linguistic disciplines, which focus exclusively on the content or the argumentative structures of speeches, without considering the relevant role of voice characteristics in the success or effectiveness of a speaker. In fact, this often happens in politolinguistics and this paper presents a possible alternative path for transcription of political speeches with BAS web tools, based on previous studies carried out for transcription in other disciplines. The use of such tools opens the perspective of including the signal in the analyses. The corpus of speeches used as examples was collected for a doctoral project on political speeches in Germany and Italy. The aim of this study is not to outline a technical perspective on the development and commands of transcription tools or software, but to show how to solve common problems that researchers in political language could face, once they want to consider for instance longer speeches delivered at political conventions and not transcribed before. Based on the problems and solutions that have been found for the doctoral project, researchers in the (polito)linguistic field could also determine whether the presented tools are applicable for their research and perhaps useful to make time-consuming tasks as transcribing easier. The article will present such tools, discuss their advantages and disadvantages, and conclude with examples and recommendation of best practices.

**Keywords:** political language; transcription; web tools; politolinguistics; audio signal.

## 1. Einleitung

Dieser Beitrag befasst sich mit den sogenannten BAS Web Tools<sup>1</sup>, die bei der Transkription bzw. Korrektur von Transkripten<sup>2</sup> und der Analyse von Reden eine wertvolle Hilfe darstellen können. Anhand von Beispielen aus einem

<sup>1</sup> <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface> (25.04.2023).

<sup>2</sup> Der Terminus ‚Transkription‘ bezeichnet den Prozess, wohingegen das ‚Transkript‘ das Ergebnis der Transkription ist.

deutsch-italienischen Korpus politischer Reden werden Wege vorgestellt, die den Prozess des Transkribierens und die Analyse des Audiosignals erleichtern. Der Beitrag basiert auf den konkreten Erfahrungen mit der Transkription für ein Korpus politischer Reden im Rahmen eines Doktorats und hat nicht die Absicht technische Anleitungen zu liefern. Das Ziel ist vielmehr, einige Tools vorzustellen, die für ForscherInnen aus der (Polito)Linguistik interessant sein können (für eine ausführliche Definition der Disziplin siehe u.a. Burkhardt 1996; Niehr 2014). In einer Pilotstudie konnte Christoph Draxler (PD am Institut für Phonetik und Sprachverarbeitung an der Ludwig-Maximilians-Universität in München) kritische Aspekte des Transkriptionsprozesses, zum Beispiel die Wortfehlerrate und die Transkriptionsgeschwindigkeit, auf der Grundlage der hier untersuchten Reden herausfiltern und analysieren (Draxler 2023). Allerdings sind die Ergebnisse noch vorläufig und eine weitere Erforschung des Themas ist wünschenswert.

Aufgebaut ist der Artikel wie folgt: Nach einer theoretischen Einführung zum Thema Transkription in sprachwissenschaftlichen Disziplinen wie der Politolinguistik (Abschnitt 2) werden ausgewählte BAS Web Tools vorgestellt (Abschnitt 3). Anhand des für das Promotionsprojekt erstellten Korpus wird eine praktische Anwendung der BAS Tools gezeigt (Abschnitt 4), bevor abschließend (Abschnitt 5) ein Fazit gezogen wird.

Das Thema des Beitrags knüpft folglich an die Sektion dieser Sonderausgabe an, die sich mit den digitalen Transformationen in diversen linguistischen Bereichen und den daraus resultierenden Implikationen beschäftigt. Webbasierte BAS Tools stellen eine neuartige digitale Möglichkeit für die Fachbereiche dar, die Transkripte verwenden, gleichzeitig bringen sie auch Veränderungen mit sich und erfordern deshalb eine auf den neuesten Stand gebrachte Expertise. Außerdem bringt dieser Ansatz eine neue Perspektive mit sich und das nicht nur im Bereich der politischen Sprache. Die Inhalte und die rhetorischen bzw. argumentativen Strukturen der Reden sind grundlegend für die Analyse der politischen Sprache; aber ebenso sind stimmliche Merkmale für die Wirkung einer Rede verantwortlich. Die Auseinandersetzung mit den Analyse-Tools bereichert die Forschung in diesem Bereich um ein weiteres brauchbares Instrumentarium.

## **2. Transkription und Segmentation von Reden**

### **2.1. Transkriptionsmethoden: ein Exkurs**

Der Begriff ‚Transkription‘ bezieht sich im Bereich der Linguistik im Allgemeinen auf zwei verschiedene Textprodukte. Zum einen versteht man unter Transkription die phonetische oder phonologische Verschriftlichung der gesprochenen Sprache unter Anwendung spezieller Alphabete, z.B. IPA oder

SAMPA. Auf der anderen Seite kann mit dem Begriff ‚Transkription‘ auch die orthographische Übertragung der gesprochenen Sprache gemeint sein. Diese findet ihre Anwendung i.d.R. in Disziplinen wie der Soziolinguistik, der Gesprächsanalyse bzw. anderen Fachgebieten, die sich vor allem mit dem Inhalt der gesprochenen Sprache beschäftigen.

Dieser Beitrag befasst sich mit der zuletzt genannten Bedeutung von Transkription, d.h. der orthographisch transkribierten gesprochenen Sprache. Die Transkription von Reden oder Gesprächen ist also ein Mittel, um z.B. Inhalte oder Interaktionen analysieren zu können. Das führt dazu, dass das Transkribieren oft als eine rein zweckmäßige und zeitaufwändige Aufgabe wahrgenommen wird, die nichts mit der eigentlichen Analyse zu tun hat, sondern nur ein Mittel zum Zweck ist. Das heißt, dass Transkripte normalerweise nicht Bestandteil der Analyse selbst sind, sondern sozusagen eine Voraussetzung, um das auditive Material sammeln und analysieren zu können. Transkripte sind aus folgenden Gründen notwendig: “necessary for investigations, for storage, reproduction, search, exemplary and presentation purposes, for it is only on the basis of a transcription that those tasks become executable” (Kirk, Andersen 2016, S. 296).

Wie erfolgt das Transkribieren zu wissenschaftlichen Zwecken üblicherweise? Da es sehr zeitintensiv ist, werden gern kurze Reden und Gespräche ausgewählt, oder es werden einfach Reden gewählt, deren Transkripte schon zur Verfügung stehen. Da jedoch die Dauer ein wichtiger Faktor dafür ist, ob eine Rede oder ein Gespräch für die Analyse ausgewählt wird, was besonders in der Politik eine wichtige Rolle spielt, bedeutet das eine Einschränkung: Es kann dazu führen, dass Reden allein aufgrund ihrer Länge von einer Analyse ausgeschlossen werden. Als weitere zeitsparende Strategie kann man sich auch auf relevante Ausschnitte einer Rede beschränken. Dabei besteht allerdings das Risiko, etwas zu übersehen, was die anschließende Analyse beeinflussen kann. Das ist insbesondere dann der Fall, wenn nur die transkribierten Teile der Rede wiedergegeben werden.

Auch bei Studien über die politische Sprache lässt sich feststellen, dass der Fokus normalerweise entweder auf schriftlichen Texten oder auf im Voraus transkribierten bzw. kurzen Reden liegt. Das Thema des Transkribierens an sich wird hier nicht unmittelbar vertieft, was die Relevanz dieses Beitrags deutlich macht. Die Untersuchungsbeispiele stammen nämlich üblicherweise von schriftsprachlichen Dokumenten oder aus Reden, für die offizielle Transkripte zur Verfügung stehen. So werden beispielsweise Wahlkampfreden, die meist vor größeren Menschenmengen gehalten werden, eher selten transkribiert. Der Grund dafür ist einerseits, dass diese Reden im Normalfall nicht im Voraus schriftlich vorliegen und andererseits, dass das Audio-Format dieser Art von Reden wegen des großen Publikums häufig Hintergrundgeräusche aufweist, die eine Transkription erschweren.

Brambilla (2007, S. 63) verweist darauf, dass es im Hinblick auf die Analyse der gesprochenen Sprache in politischen Reden eine Forschungslücke gibt, obwohl die stimmliche Umsetzung vor allem in Bezug auf die Persuasion die Hauptrolle spielt. Außerdem hebt Brambilla hervor, dass offizielle Transkripte politischer Reden oft von Stenografen geglättet wurden und deswegen keine zuverlässige Quelle für die Analyse darstellen (vgl. ebd., S. 66). Das heißt, dass die Transkriptionsprobleme nicht nur Reden betreffen, die noch nie transkribiert wurden, sondern auch diejenigen, für die offizielle Transkripte existieren. Beispielsweise stammt das offizielle Transkript einer von Brambilla analysierten 40-minütigen politischen Rede (ebd., S. 111-126) auch von Stenografen. Wenn man die Rede nämlich auf YouTube abspielt, merkt man sofort, dass die Transkription nicht mit dem gesprochenen Text übereinstimmt. Das zeigt, dass ein zuverlässigeres Transkript entstehen würde, wenn sich der Zeitaufwand reduzieren ließe und man die Transkription computergestützt durchführen könnte. Zudem zeigt es, dass Transkriptionshindernisse nicht von den Textsorten bzw. Textarten abhängig sind, sondern verschiedene Textsorten gleichermaßen betreffen.

Das Transkribieren ist nämlich ein komplexer Prozess und daher haben sich viele verschiedene Forschungsansätze dazu entwickelt, die bis heute noch keinen wirklichen gemeinsamen Nenner haben. Fuß und Karbach (2019) unterscheiden beispielsweise zahlreiche Transkriptionstypen auf der Basis der Transkriptionsregeln und der Phänomene, die verschriftlicht werden sollten (ebd., S. 20). Sie stellen fest, dass „das Transkript [...] letztendlich als das Ergebnis einer Transformation von Daten zu wissenschaftlichen Zwecken [ist]“ (ebd., S. 28). Aus diesem Grund spielt es auf jeden Fall eine Rolle, wie die Transkription gestaltet und durchgeführt wird.

Heilmann (2003) beschreibt drei verschiedene Transkriptionsmethoden und stellt dabei die Schwierigkeiten des Transkriptionsprozesses dar. Laut Heilmann (ebd., S. 142) dienen Transkriptionen dazu, „Mündlichkeit zu verschriftlichen“, wobei Parameter festgesetzt werden müssen, die eine Analyse möglich machen. In Bezug auf den Transkriptionsprozess erkennt Heilmann (ebd.) verschiedene Paradoxien des Transkribierens. Eine davon ist die „Tatsache, dass [...] die transkribierten Äußerungen in ihrer situationsgebundenen Mündlichkeit so weit als nur irgend möglich erhalten bleiben sollen“. Es scheint in der Tat widersprüchlich, die gesprochene Sprache zu verschriftlichen, wobei man gleichzeitig angehalten werde, darauf zu achten, die Kennzeichen der Mündlichkeit bei den Transkripten nicht zu verlieren. Je nach Forschungszweck der Transkripte existieren verschiedene Transkriptionsprototypen. Heilmann fokussiert sich auf drei davon (ebd., S. 144-150).

Der erste Typ ist die *integrierte Zeilenschreibung*, die laut Heilmann (ebd., S. 144) das GAT (Gesprächsanalytisches Transkriptionssystem, online

verfügbar)<sup>3</sup> als Notationssystem verwendet. Solch ein Notationstyp benötigt keine besondere Software, vielmehr wird i.d.R. ein simples Textdokument verfasst. Obwohl Pausen und Interjektionen auch notiert werden, zielt diese Art von Transkripten darauf ab, die „textuelle Verflechtung“ (ebd.) zu verdeutlichen. Seit 2003, d.h. seit Veröffentlichung des eben zitierten Beitrags von Heilmann, hat sich das GAT-System verändert. Im Jahr 2009 wurde nämlich GAT 2 eingeführt (Selting *et al.* 2009), das der internationalen Forschungsgemeinschaft dienen soll und bei dem Transkriptionskonventionen weiterentwickelt wurden (ebd., S. 354). Heilmann (2003, S. 144) weist darauf hin, dass bei der Verwendung von GAT und seiner integrierten Zeilenschreibung aufgrund eindeutiger Konventionen Fehler reduziert und Irrtümer verhindert werden.

Die anderen beiden von Heilmann (ebd.) dargelegten Transkriptionsprototypen sind die *Partiturschreibung* und die *Zeitreihennotation*. Bei der Partiturschreibung

[werden] die unterschiedlichen Ebenen gesprochener Sprache (linguistische, paralinguistische und extralinguistische) [...] – einer Partitur vergleichbar – übereinander angeordnet, so dass eine synchronoptische Darstellung entsteht. (Heilmann 2003, S. 145)

Normalerweise wird so eine Art von Transkription von einer Software unterstützt. Das dazugehörige Notationssystem ist laut Heilmann (vgl. ebd., S. 145) HIAT (Halbinterpretatives Arbeitstranskript, online verfügbar).<sup>4</sup> Während man den Umgang mit der Software erlernen muss, benötigt auch das Transkribieren relativ viel Zeit, insbesondere wenn die Transkription sehr ausführlich sein soll. Ein kostenloses häufig benutztes Programm dafür ist EXMARALDA,<sup>5</sup> das heruntergeladen und ebenfalls regelmäßig aktualisiert werden muss.

Die *Zeitreihennotation* (Heilmann 2003, S. 148-149) ist schließlich ein visueller Transkriptionsprototyp, bei dem „in Analogie zur gesprochenen Sprache als Folge einzelner Laute [...] das menschliche Bewegungsverhalten als Folge von Positionszuständen einzelner Körperteile durch Ziffernfolgen abgebildet“ wird (ebd., S. 148). Diese Art von Transkription ist besonders zeitaufwändig und man braucht eine gewisse Expertise, um sie durchführen zu können. Es handelt sich eher um eine Art von Visualisierung, welche die Körpersprache einbezieht (z.B. Blickverhalten, Gestik) und mit der Verschriftlichung des gesprochenen Textes verbindet.

Heilmann (vgl. ebd., S. 150) erläutert dazu, dass die ausgewählte Transkriptionsmethode vom Untersuchungsziel abhängt. Alle drei von

<sup>3</sup> <https://agd.ids-mannheim.de/gat.shtml> (04.06.2023).

<sup>4</sup> [https://www.exmaralda.org/hiat/files/azm\\_56.pdf](https://www.exmaralda.org/hiat/files/azm_56.pdf) (04.06.2023).

<sup>5</sup> <https://exmaralda.org/en/> (25.04.2023).

Heilmann dargestellten Methoden weisen gewisse Stärken und Schwächen<sup>6</sup> auf und setzen zudem eine manuelle Transkription voraus.

Das Projekt FOLK (Forschungs- und Lehrkorpus Gesprochenes Deutsch)<sup>7</sup> bietet die Möglichkeit, ausgewählte transkriptionsbezogene Aspekte zu untersuchen. Allerdings stammt die gesprochene Sprache in FOLK hauptsächlich aus Interaktionen (Schmidt 2016, S. 397), weshalb ein Transkriptionsinstrumentarium ausgewählt wurde, das dafür geeignet ist. Interaktionstranskriptionen stellen weiterhin eine Herausforderung für die in diesem Beitrag präsentierten Tools dar, was in den folgenden Abschnitten genauer erläutert wird. Schmidt (ebd., S. 404) hebt vor allem den Zeitaufwand für die manuelle Transkription hervor. Er schlägt jedoch nicht vor, die automatische Spracherkennung (ASR) zu verwenden, sondern den Fokus auf die uniformierten Transkriptionskonventionen und ein benutzerfreundliches Tool zu legen, das leicht zu erlernen und zu verwenden ist (vgl. ebd., S. 404-405). Daher wurde die Software FOLKER<sup>8</sup> ausgewählt, bei der der Text automatisch auf GAT-Konventionen überprüft wird und jedem/r SprecherIn eine Transkriptionsspur zugewiesen werden kann (ebd., S. 407). Laut Schmidt (ebd., S. 413) kann FOLKER den Transkriptionsprozess darüber hinaus optimieren und beschleunigen, obwohl es sich immer noch um eine manuelle Transkription handelt. Im Abschnitt 3 wird auch darüber reflektiert, ob und wie die vorgestellten Tools andere Programme ergänzen können. Es wäre z.B. gut möglich, FOLKER und Web Tools zusammen zu verwenden.

Wenn auch die oben beschriebenen Methoden gültige Transkriptionsmethoden sind, sollte man auch andere Methoden kennen, um sich je nach Forschungsziel die richtige Software bzw. das richtige Tool auszusuchen.

## **2.2. Pausenberechnung und Pausenmarkierung**

Wie schon erwähnt wurde, ist es schwierig, die genaue Dauer von Pausen wiederzugeben. Notationssysteme erlauben es normalerweise, die Dauer der Pausen anzugeben, wobei die Eingabe manuell gemacht werden muss. Dies gilt für alle oben genannten Transkriptionsprototypen.

Für die manuelle Markierung bieten Fuß und Karbach (2019, S. 44) folgende Visualisierungsmöglichkeiten. Eine Möglichkeit besteht darin, Pausen sekundengenau zu transkribieren, während eine andere Option darin besteht, die ungefähre Pausenlänge in drei definierten Längen grob zu unterscheiden. Demzufolge bezeichnet „(..)“ eine Pause von zwei Sekunden

<sup>6</sup> Zum Beispiel kann die Dauer der Pausen fehlerhaft transkribiert werden, wie im folgenden Abschnitt vertieft wird.

<sup>7</sup> [https://agd.ids-mannheim.de/FOLK\\_extern.shtml](https://agd.ids-mannheim.de/FOLK_extern.shtml) (07.10.2023).

<sup>8</sup> <https://agd.ids-mannheim.de/folker.shtml> (07.10.2023).

und „(...)“ eine Pause von bis zu fünf Sekunden, während längere Pausen als „(Pause)“ vermerkt werden (ebd.). FOLKER würde es ermöglichen, Pausen automatisch zu messen und zu transkribieren (Schmidt 2016, S. 408).

Bei den im Folgenden präsentierten BAS Web Tools besteht die Option, Pausen überhaupt nicht zu transkribieren oder sich für das automatische Anzeigen von Zeitmarkern zu entscheiden. Obwohl man auch hier auf Schwierigkeiten stoßen kann, ist die Fehleranfälligkeit bei der automatisierten Dauermessung der Pausen gering.

Schließlich gibt es die Variante, die Pausen überhaupt nicht zu markieren und sie direkt im Zusammenhang mit Klangtools zu beobachten und zu untersuchen. Dies reduziert nicht nur mögliche Fehler, sondern macht die Transkripte auch leichter lesbar. Mit diesem Ansatz kann man sowohl den Inhalt bzw. die textuelle Struktur als auch das Signal untersuchen.

Man könnte meinen, dass das Thema der Pausenproblematik nebensächlich sei, aber in linguistischen Disziplinen spielt es eine grundlegende Rolle. Sowohl in Interaktionen als auch in Reden sind die Pausen und ihre Dauer von großer Bedeutung. Aus diesem Grund sollten sie so präzise wie möglich berechnet und angegeben werden. Das Ergebnis von den hier präsentierten BAS Web Tools ist daher ein zeitaligniertes Transkript, mittels dessen man die verschiedenen Phänomene beobachten kann. Zudem kann man wissen, wann sie stattfinden. Die Tools verfügen dazu über verschiedene Export-Möglichkeiten in passenden Formaten. Dabei bieten sie Funktionen an, die es ermöglichen, den Wechsel zwischen Klang und Pausen zu untersuchen.

### 3. BAS Web Tools

BAS ist die Abkürzung für das Bayerische Archiv für Sprachsignale. Es wurde im Jahr 1995 gegründet und gehört seit 2013 zur CLARIN Infrastruktur.<sup>9</sup> Das BAS umfasst verschiedene Tools für Phonetik und Sprachverarbeitung und beinhaltet auch genau die Tools, die in diesem Beitrag vorgestellt werden (vgl. Kisler *et al.* 2017). OCTRA und WebMAUS werden ausführlich präsentiert, während auf EMU-webApp nur kurz hingewiesen wird.

Die BAS Web Tools können entweder auf dem Browser benutzt werden (sie sind kostenfrei zugänglich mittels einer E-Mail-Adresse, die der CLARIN-Infrastruktur zugehörig ist) oder direkt in der Programmiersprache R<sup>10</sup> durch das Paket emuR<sup>11</sup> aufgerufen werden (Poerner, Winkelmann 2018, S. 141): „the BAS-emuR interface, which allows users to apply BAS web services to

<sup>9</sup> <https://www.bas.uni-muenchen.de/Bas/BasHomedeu.html> (25.04.2023).

<sup>10</sup> R ist eine Programmiersprache für statistische Berechnungen und Grafiken: <https://www.r-project.org/> (09.02.2024).

<sup>11</sup> R-Pakete sind Erweiterungen für die Programmiersprache R. Sie enthalten Codes, Befehle und Daten. emuR: <https://cran.r-project.org/web/packages/emuR/index.html> (09.02.2024).

their EMU databases from within the R environment“. Laut Poerner und Winkelmann (ebd., S. 143) liegt der größte Vorteil dieser Option darin, dass man direkt „[a] ready-made EMU database“ bekommt, was heißt, dass man keine Formate konvertieren muss (ebd.). Ein Nachteil, der genannt wird, ist, dass man eine stabile Internetverbindung braucht. Des Weiteren müssen die Dateien temporär auf dem BAS gespeichert werden (ebd.). Man muss also die Programmiersprache R und das Paket emuR gut beherrschen, um diese Anwendung nutzen zu können. Wenn man die große Palette an BAS Tools zum ersten Mal verwenden möchte, ist es unwahrscheinlich, dass man emuR schon kennt. Aus diesem Grund wurde in diesem Beitrag beschlossen, die Tools einzeln zu besprechen, wenn sie auch mittels emuR auf R aufgerufen werden könnten. Dies ist u.a. deshalb sinnvoll, weil diese Tools auch von LinguistInnen aus Bereichen außerhalb der Phonetik verwendet werden können.

BAS Web Tools unterscheiden sich von Transkriptionsprogrammen, die als Software entwickelt wurden, d.h. die auf dem Betriebssystem eines Geräts heruntergeladen werden und nicht direkt auf einem Server laufen, wie z.B. EXMARaLDA und ELAN.<sup>12</sup> Die unterschiedlichen Leistungen der Betriebssysteme beeinflussen nämlich auch die Leistung der Software. Die webbasierten Tools sind hingegen von der Leistung des jeweiligen Betriebssystems unabhängig, weil die Verarbeitung der Dateien auf dem Server erfolgt (vgl. Kisler *et al.* 2016, S. 3882), d.h. dass man nur ein Gerät und eine stabile Internetverbindung dazu braucht. Die Einstellungen der verwendeten Sprachen können leicht geändert werden, da man sie einfach online auswählen kann, ohne dass die Installierungssprache mit der jeweiligen Software interferiert (vgl. ebd.). Die webbasierten Tools wurden gerade deshalb entwickelt, um Schwierigkeiten bei der Verwendung lokaler Software zu umgehen. Das bedeutet, dass softwarebedingte Probleme umgangen bzw. ausgeschlossen werden können.

Gelegentlich kann es beim Hochladen großer Dateien auch passieren, dass die Software überfordert ist und es sogar zu einem Systemabsturz kommt. Dies ist bei den webbasierten Tools nicht zu erwarten:

This allows access for multiple users at the same time without slowing down the processing for other users. All web interfaces that allow to upload multiple files per drag and drop make use of this feature and process, for one user, multiple files at the same time. (Kisler *et al.* 2016, S. 3882)

Mit BAS webbasierten Tools ist es möglich, eine Verknüpfung unterschiedlicher Programme, die nicht webbasiert sind, herzustellen (siehe Abschnitt 2.1). Wie Abbildung 1 zeigt, kann man beispielsweise die (externe)

<sup>12</sup> <https://archive.mpi.nl/tla/elan> (25.04.2023).

Software EXMARaLDA mit WebMAUS, das zu den BAS Web Tools gehört, verbinden.

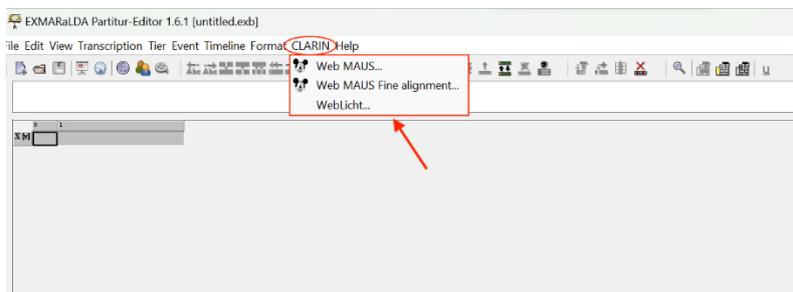


Abbildung 1

Zugang zu BAS Web Services durch EXMARaLDA (siehe Markierung).

Auch weitere web- und softwarebasierte Programme können durch den Export in beliebig viele Dateiformate verwandelt werden.

Außerdem sind die BAS Web Tools benutzerfreundlich und wurden im Laufe der Zeit für zahlreiche nicht phonetische Disziplinen (z.B. *Oral History*) weiterentwickelt (vgl. Pömp, Draxler 2017, S. 145). Somit stellt sich auch die Frage, ob die Tools, die für phonetische Fragestellungen und Sprachverarbeitung entwickelt wurden, auch für andere sprachwissenschaftliche Disziplinen wie die Politolinguistik geeignet sind. Im Folgenden werden zwei ausgewählte Tools vorgestellt: OCTRA<sup>13</sup> und WebMAUS General<sup>14</sup>.

### 3.1. OCTRA

OCTRA (*OrthographiC TRAnscription*) ist ein Tool, das die orthographische Transkription von Audio Files ermöglicht. Es wird einfach mit dem Browser verwendet, d.h. dass man keine Software herunterladen muss. Außerdem bietet OCTRA zahlreiche verschiedene Output-Formate, damit man die Transkripte mit der jeweils gewählten Software und anderen Tools weiter bearbeiten kann. Das Handbuch für OCTRA<sup>15</sup> steht online zur Verfügung, so dass man sich einerseits über die Funktionen und andererseits über die Ziele dieses Tools informieren kann. Der Ablauf des Transkribierens mit OCTRA wird in einzelnen Schritten anhand von Beispielen und Bildern erläutert.

Was die Konventionen für die Transkription angeht, braucht man normalerweise einige Regeln. Bei OCTRA existieren allerdings keine vorgegebenen Regeln wie bei GAT, vielmehr können die transkribierenden Personen ihre eigenen Richtlinien definieren oder entsprechende Icons

<sup>13</sup> <https://clarin.phonetik.uni-muenchen.de/apps/octra/octra/login> (16.05.2023).

<sup>14</sup> <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSGeneral> (25.04.2023).

<sup>15</sup> <https://clarin.phonetik.uni-muenchen.de/apps/octra/manual/1.4.0/> (25.04.2023).

auswählen, wie Abbildung 2 zeigt. Dies funktioniert, solange die Konventionen in den Transkripten konsistent sind.



Abbildung 2  
Icons auf OCTRA-Transkriptionskonsole (siehe Markierung).

Was den Prozess der Transkriptionen bei OCTRA angeht, merken Pömp und Draxler (2017, S. 146) Folgendes an:

OCTRA (for *OrthographiC TRAnscription*) is based on the concept of time-aligned transcription chunks: transcribers insert segment boundaries at appropriate positions in the signal, usually within pauses or at the end of intonation phrase.

In OCTRA schreibt man unabhängig von der Anzahl der SprecherInnen auf einer einzigen Spur. Die verschiedenen InterlokutorInnen können unterschieden werden, indem man ein Zeichen setzt, wenn eine andere Person zu sprechen beginnt. Üblich ist die Kennzeichnung der einzelnen SprecherInnen mit <S0,1,..>. Es gibt zwar die Möglichkeit, weitere Spuren hinzuzufügen, dies ist allerdings nur im lokalen Modus möglich (siehe Abbildung 3). Man sollte beachten, dass es nicht möglich ist, die verschiedenen Spuren untereinander und somit gleichzeitig zu betrachten. Wenn man Gespräche visualisieren möchte, dann sollte man Konventionen einsetzen, die es ermöglichen, beispielsweise Überlappungen und andere typische Elemente auf nur einer Spur zu zeigen. So ist eine Möglichkeit, Überlappungen mit dem Zeichen <ovl> zu markieren. Kirk & Andersen (2016, S. 296) verweisen in diesem Zusammenhang auf den organisatorischen Vorteil eines aus mehreren Spuren bestehenden Transkripts, was die Übersichtlichkeit, Klarheit und Lesbarkeit angeht. Das heißt, dass man sich bei der Entscheidung für das passende Programm von den Forschungszwecken seines Projekts leiten lassen soll. Dabei sollten sowohl technologische als auch methodologische Aspekte berücksichtigt werden (Schmidt 2016, S. 401).

Wenn man OCTRA öffnet, erscheinen zwei Modi (Abbildung 3): der Online-Modus (links) und der lokale Modus (rechts). Der Online-Modus eignet sich ein Projekt, das nur für seine Mitglieder zugänglich ist. Beim lokalen

Modus kann man autonom Audio Files hochladen und die Transkription anfangen.

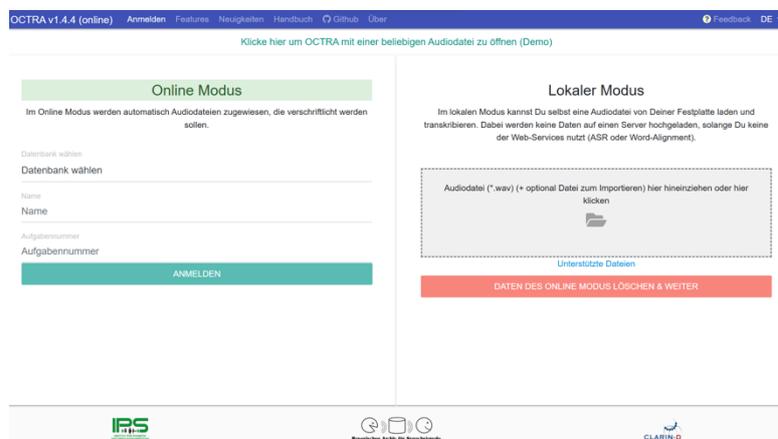


Abbildung 3  
Online-Modus (links) und lokaler Modus (rechts) auf OCTRA.

Transkriptionseinheiten können unterschiedlich lang sein und sie werden beliebig aufgeteilt. Damit der Redefluss nicht unterbrochen wird, werden die Grenzmarker üblicherweise bei Pausen platziert (Abbildung 4).



Abbildung 4  
Trennungszeichenplatzierung auf OCTRA (siehe Markierung).

Danach kann man auf die jeweilige Einheit klicken und ein Segment nach dem anderen transkribieren (Abbildung 5).

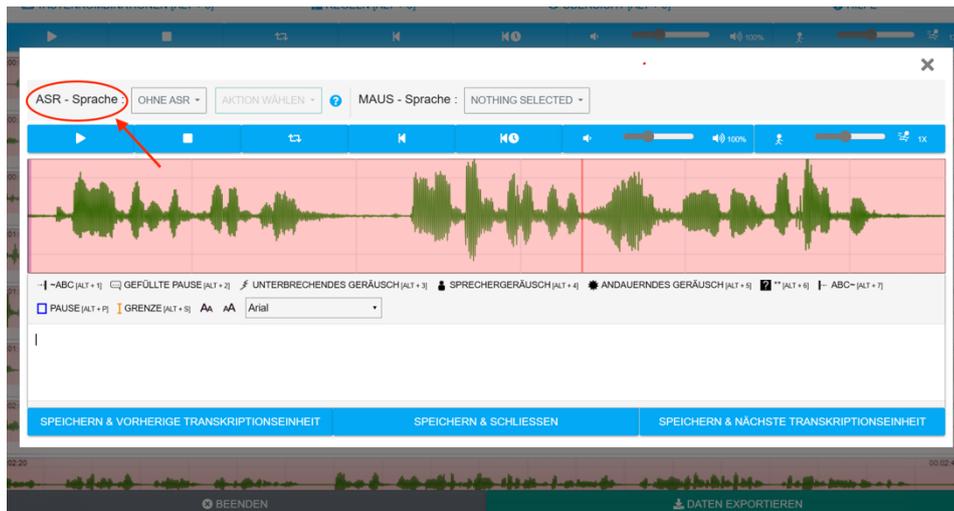


Abbildung 5

Überblick über ein einzelnes zu transkribierendes Segment und die ASR in OCTRA.

In der Abbildung 5 oben links gibt es die Möglichkeit, die automatische Spracherkennung (ASR) zu nutzen und sie entweder nur auf die geöffnete Einheit oder auf den ganzen Text anzuwenden. Die Anwendung der automatischen Spracherkennung wird im Abschnitt 4 vertieft.

Nach Beendigung der Transkription kann man diese mit einem Klick auf „Übersicht“ anschauen. Das Transkript ist schließlich in verschiedenen Formaten exportierbar, u.a. im txt-Format, das beispielsweise für WebMAUS benötigt wird (siehe Abbildung 6).

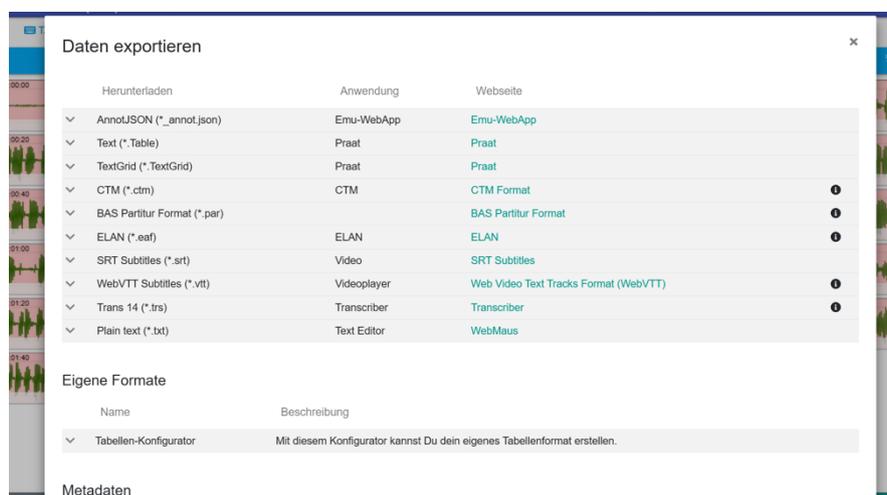


Abbildung 6

Tabelle der Output-Formate in OCTRA.

Die auf diese Weise erstellten Texte enthalten die orthographische Transkription und können für alle möglichen linguistischen Analysen verwendet werden. Innerhalb des txt-Formates können verschiedene Exportvarianten ausgewählt werden. Beispielsweise kann man entscheiden, ob

die Zeitangaben in den Text integriert werden sollen, oder ob der Text ohne Zeitangaben gespeichert werden soll. Beispiele von Transkripten mit und ohne Zeitangaben aus den italienischen und deutschen politischen Reden sind im Abschnitt 4 zu finden.

### 3.2. WebMAUS

WebMAUS (*Munich AUtomatic Segmentation*) ist ein weiteres BAS Tool und wird oft im Zusammenhang mit OCTRA benutzt. Dieses Tool ist hilfreich für unterschiedliche Anwendungen, z.B. für Redeanalysen auch im Bereich der (Polito)Linguistik, weil es, wie bereits erwähnt, die Möglichkeit bietet, Signale zu analysieren. Wie im Abschnitt 4 dargelegt wird, können zunächst OCTRA und anschließend WebMAUS eingesetzt werden. WebMAUS bietet die automatisierte Segmentierung bzw. Alignierung von Audiosignalen und den jeweils zugehörigen Transkripten<sup>16</sup> nicht nur für die Segmente, die mit OCTRA bearbeitet wurden, sondern auch für einzelne Laute: Es können also sowohl die Position jedes Wortes im Signal bestimmt als auch die Pausen präzise gemessen werden. Um eine Analyse durchzuführen, die auf der einen Seite das Signal und auf der anderen textuelle Bestandteile berücksichtigt, ist WebMAUS bestens geeignet. Außerdem müssen Pausen in den Transkripten, wie bereits erwähnt, nicht unbedingt geschrieben werden, da man sie in WebMAUS direkt visualisieren kann. Insbesondere kann man je nach Forschungszweck verschiedene Output-Transkripte einrichten: Wenn man z.B. lediglich die Wortwahl oder die Syntax eines politischen Textes analysieren will, spielen die Pausen eine untergeordnete Rolle. Ganz davon abgesehen, dass die Transkripte ohne Pausenmarkierung leichter zu lesen sind. Wenn man dagegen den Redefluss untersuchen will, kann man dies mit den mit Hilfe von WebMAUS verarbeiteten Files im Detail untersuchen, da der Text mit der Sprachaufnahme aligniert wird.

Grundsätzlich gibt es drei Varianten von WebMAUS, die zur Verfügung stehen: WebMAUS Basic,<sup>17</sup> WebMAUS General<sup>18</sup> und Pipeline with(out) ASR.<sup>19</sup> WebMAUS Basic (Abbildung 7) ist, wie der Name vermuten lässt, die Basisversion von WebMAUS und enthält weniger Funktionen als die erweiterte Version WebMAUS General.

<sup>16</sup> <https://www.clarin.eu/showcase/webmaus-automatic-segmentation-and-labelling-speech-signals-over-web> (11.10.2023).

<sup>17</sup> <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic> (25.04.2023).

<sup>18</sup> <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSGeneral> (25.04.2023).

<sup>19</sup> <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/Pipeline> (25.04.2023).

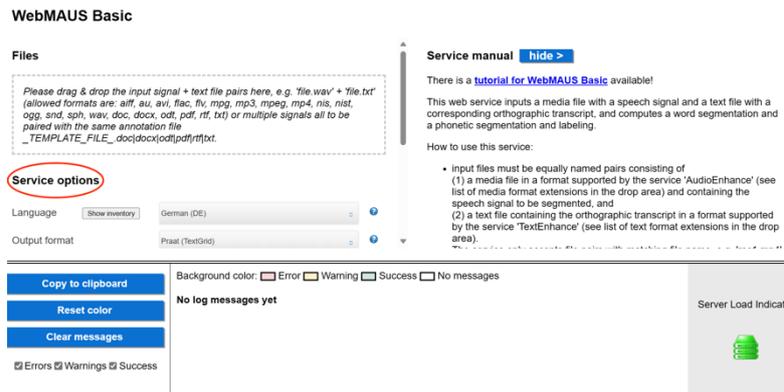


Abbildung 7

Homepage von WebMAUS Basic und Überblick über die Funktionen (siehe Markierung).

Abbildung 7 zeigt die Startseite der Browserversion von WebMAUS Basic, auf der bereits ersichtlich ist, dass nur wenige Einstellungen zur Verfügung stehen. Man muss jeweils eine Audiodatei im wav-Format und eine gleichnamige Text-Datei im txt-Format hochladen, um das Programm zu starten. Daneben kann man die gewünschte Sprache und ein passendes Output-Format auswählen. Im unteren Feld werden hingegen Status- sowie Fehlermeldungen angezeigt. Nachdem die Dateien verarbeitet und exportiert wurden, kann man sie herunterladen. Für größere Dateien wird ein komprimierter ZIP-Ordner automatisch erstellt. Leider besteht in WebMAUS Basic nicht die Möglichkeit, sich per E-Mail über den Status der Verarbeitung benachrichtigen zu lassen (dies gehört zu den *Expert Options*), selbst wenn die Anzahl der zu bearbeitenden Dateien groß ist und der Export lange dauert.

Das letztgenannte Problem wurde von WebMAUS General gelöst. In dieser fortgeschrittenen Version von WebMAUS Basic gibt es eine Reihe erweiterter Einstellungen (*Expert Options*), wie die Abbildungen 8-11 darstellen:



Abbildung 8

*Service Options* von WebMAUS General.

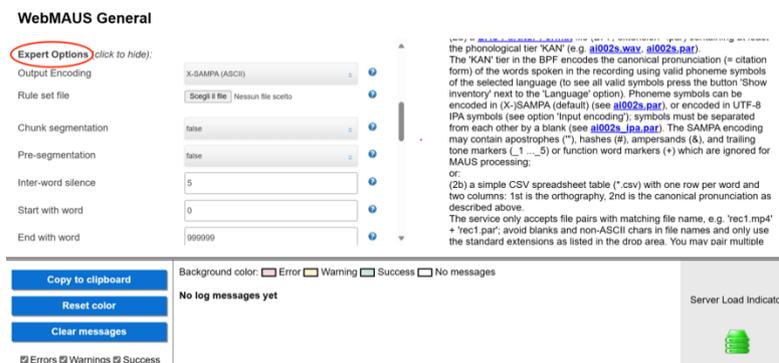


Abbildung 9  
*Expert Options* von WebMAUS General (siehe Markierung).



Abbildung 10  
*Weitere Expert Options* von WebMAUS General.

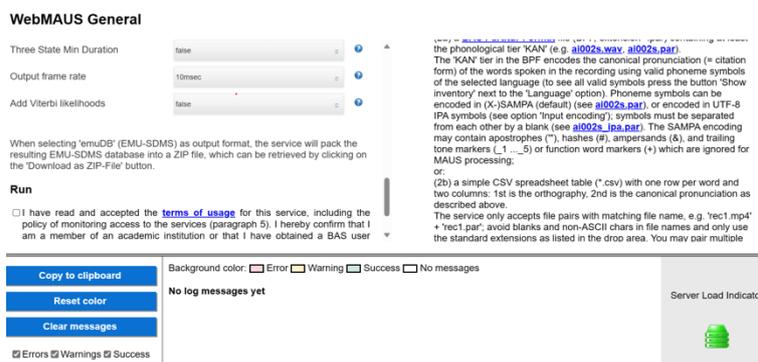


Abbildung 11  
*Weitere Expert Options* von WebMAUS General.

Unter diesen Optionen gibt es auch die Möglichkeit, sich Benachrichtigungen per E-Mail senden zu lassen. Obwohl WebMAUS General die erweiterte Version von WebMAUS Basic ist und daher zahlreiche zusätzliche Funktionen umfasst, bietet es keine Lösung für ein wichtiges Problem an, dass nämlich WebMAUS General nur Dateien im BAS Partitur Format (BPF) erkennt. Das heißt, dass diese jedes Mal zuerst konvertiert werden müssen. Zwar wurde das

BAS-Tool Annotconv<sup>20</sup> extra dafür entwickelt, aber gleichzeitig kann man damit nur BPF in andere Formate konvertieren und nicht umgekehrt. Die Verwendung von WebMAUS General könnte aufgrund dieser Formatprobleme aufwendig sein.

Um WebMAUS General dennoch nutzen zu können, ohne die Dateien vorher mittels BPF konvertieren zu müssen, wird von BAS die Anwendung Pipeline with(out) ASR als mögliche Lösung angeboten. Die Pipeline ruft WebMAUS General auf und ermöglicht auch dort, die zahlreichen erweiterten Einstellungen zu öffnen. Anschließend müssen dann, wie auch in WebMAUS, jeweils eine wav- und eine txt-Datei hochgeladen werden. Auch hier muss man wieder darauf achten, dass zusammengehörige Dateipaare denselben Namen haben. Die Abbildung 12 zeigt die Startseite der Anwendung Pipeline without ASR:

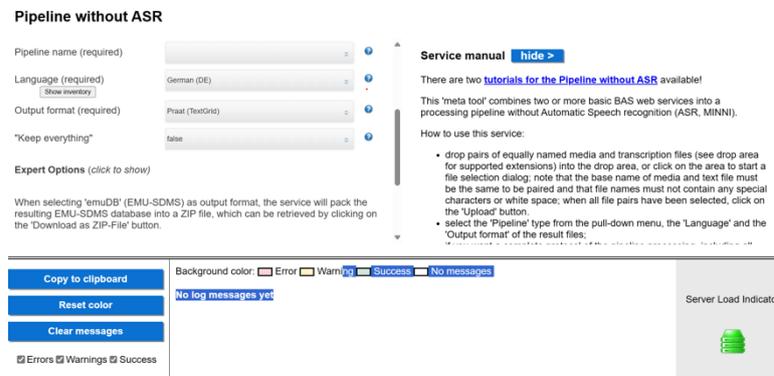


Abbildung 12  
Homepage von Pipeline without ASR.

Mit Hilfe von Pipeline kann man die von WebMAUS General bearbeiteten Dateien in einem beliebigen Output-Format erhalten und sie anschließend mit weiteren Tools oder weiteren Softwareprogrammen analysieren. Das Programm verfügt über zahlreiche Output-Formate und verfügbare Sprachen, wodurch die Weiterverarbeitung mit weiteren Anwendungen erleichtert wird. Für das erstellte Korpus werden sowohl TextGrid als auch annotJSON als Output-Formate benutzt, weitere Formate die zu anderen Applikationen passen, wie z.B. ELAN, können gewählt werden. Des Weiteren kann die Zahl der Spuren, die man mit WebMAUS bearbeiten möchte, festgelegt werden. Es wurde bereits angemerkt, dass die Pausen auf OCTRA nicht unbedingt annotiert werden müssen, da WebMAUS sie bearbeitet. Das Signal der Reden kann dann zusätzlich beispielweise auf Praat<sup>21</sup> oder auf EMU-webApp<sup>22</sup> angezeigt und analysiert werden (Abschnitt 4).

<sup>20</sup> <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/AnnotConv> (25. 04. 2023).

<sup>21</sup> <https://www.fon.hum.uva.nl/praat/> (12.10.2023).

<sup>22</sup> <https://ips-lmu.github.io/EMU-webApp/> (25.04.2023).

### **3.3. Potenzial und Herausforderungen der Verwendung von BAS Web Services**

Die BAS Web Tools bieten diverse Verwendungsmöglichkeiten in unterschiedlichen Disziplinen, so auch in der (Polito)Linguistik. Auf einige Vorteile und potenzielle Probleme bzw. Hürden soll abschließend zusammenfassend eingegangen werden.

#### *3.3.1. Potenzielle Vorteile von BAS Web Tools*

Aus der Sicht der BenutzerInnen ist einer der wichtigsten Vorteile zweifellos, dass man kein Programm auf dem Rechner installieren muss. Installierte Programme sind im Allgemeinen oft problematisch, da es nicht immer möglich ist, die neueste und gleichzeitig die für das eigene Betriebssystem kompatible Version herunterzuladen. Diese Problematik erscheint im ersten Moment oft unwichtig, doch wenn man für ein Forschungsprojekt Dateien mit geeigneter Software transkribieren und analysieren möchte und dafür viel Zeit und Mühe investiert, ist es relevant, das Risiko einer niedrigen Leistung des Betriebssystems oder des Systemabsturzes auf jeden Fall zu vermeiden.

Ein weiterer Vorteil ist, dass BAS Web Tools benutzerfreundlich sind. BenutzerInnen können zahlreiche Erklärungen auf der Homepage, Handbücher und Online-Tutorials konsultieren. Da man die Verwendungssprache auswählen kann, sind die Status- und Fehlermeldungen im Benachrichtigungsfeld zudem für alle verständlich.

Dazu kommt ein weiterer Pluspunkt der BAS Web Tools, und zwar die Interoperabilität sowohl mit anderen webbasierten als auch mit nicht webbasierten Programmen. Aufgrund der Vielfalt möglicher Output-Formate kann man die Tools in Verbindung mit anderen Instrumenten oder Softwareprogrammen nutzen, insbesondere da man einige Applikationen nur mit bestimmten Formaten verarbeiten kann (u.a. Praat mit TextGrid). Dies ist für die Arbeit in Forschungsprojekten, in denen es üblich ist, mit einer Vielzahl verschiedener Programme zu arbeiten, von großer Bedeutung.

Letztendlich wird es dank OCTRA möglich, den Transkriptionsprozess in hohem Maß zu beschleunigen, insbesondere mit der ASR, die für die Transkription, auch im Bereich der politischen Sprache, eine entscheidende Rolle spielen kann (siehe Abschnitt 4). Bisher sind keine Studien darüber in der Politolinguistik bekannt, daneben fehlen auch Beispielfälle, wo die ASR-Systeme dieses Beitrags für das Transkribieren politischer Reden verwendet werden. Einerseits hat dies auf jeden Fall mit der schon dargelegten traditionellen Trennung zwischen textuellen Elementen und dem Signal zu tun, das in der Politolinguistik selten berücksichtigt wird. Andererseits ist das Potenzial dieses Verfahrens in der Politolinguistik meistens noch nicht

genügend erkannt worden, so dass dieser Beitrag dafür gedacht ist, auf diese Verwendungsmöglichkeiten hinzuweisen.

### 3.3.2. *Hürden bei Bas Web Tools*

Probleme bei der potenziellen Verwendung von BAS Web Tools können verschiedener Art sein. Die erste, vielleicht triviale Problematik ist, dass man eine stabile Internetverbindung braucht. Üblicherweise ist eine Internetverbindung mittels WLAN vielerorts verfügbar, dennoch ist diese nicht immer und überall stabil. Des Weiteren benötigt man eine E-Mail-Adresse, die mit einer Universität oder einem Forschungszentrum der CLARIN-Infrastruktur verbunden ist, um kostenlosen Zugang zu Webdiensten und Inhalten des Repository zu erhalten.

Der Umgang mit technischen Schwierigkeiten könnte eine weitere Herausforderung darstellen. Man kann zwar jederzeit den Support um Hilfe bitten und ein Feedback hinterlassen, doch selbst wenn man schnell eine Antwort bekommt, ist sie meist nicht unmittelbar.

In Bezug auf die Transkription mit OCTRA ist einer der Nachteile, dass man lediglich auf einer einzelnen Spur transkribieren kann. Wenn mehrere SprecherInnen interagieren, wäre es hilfreich, verschiedene Personen auf unterschiedlichen Spuren gleichzeitig transkribieren zu können. Immerhin verfügt OCTRA im lokalen Modus über die Option, mehrere Spuren zu benutzen, allerdings nicht gleichzeitig. Daher ist das Tool momentan nicht besonders geeignet, Interaktionen mit vielen Überlappungen und Sprecherwechseln zu transkribieren.

Das letzte Problem, das hier genannt werden soll, bezieht sich auf die Anwender-Reichweite der Tools. Die BAS Web Tools sind nicht überall bekannt. An italienischen Universitäten tendiert man z.B. dazu, lokal installierte Programme wie EXMARaLDA oder ELAN zu verwenden. Das kann selbstverständlich an eigenen Präferenzen und Forschungsbedürfnissen liegen, doch fehlt der Vergleich mit Alternativen. Der Grund, warum die Tools in verschiedenen linguistischen Disziplinen nicht so bekannt sind, ist vermutlich auch, dass sie nur in Verbindung mit der Phonetik wahrgenommen werden. Die zuletzt genannte Tatsache kann auch die geringe Reichweite in der (Polito)Linguistik erklären. Es scheint also eine Frage der Zeit zu sein, bis die Tools auch in Bereichen außerhalb der Phonetik bekannter werden. Einige der Tools sind zwar eher für die rein phonetische Analyse gedacht, doch andere, wie die hier vorgestellten OCTRA und WebMAUS, können auch in anderen linguistischen Disziplinen genutzt werden und den Prozess des Transkribierens und der Analyse erheblich erleichtern.

## 4. Web Tools in der Politolinguistik am Beispiel italienischer und deutscher politischer Reden

In diesem Abschnitt wird eine praktische Anwendung der präsentierten Tools gezeigt. Die exemplarischen Reden, die dabei Verwendung finden, stammen aus dem Korpus des genannten Promotionsprojekts.

Das Korpus besteht aus 22 italienischen und 22 deutschen politischen Reden, die zwischen 2015 und 2022 gehalten wurden. Die Hälfte der Reden stammt von Männern und die andere Hälfte von Frauen, wobei auf die Ähnlichkeit in der politischen Orientierung geachtet wurde. Die Gesamtdauer der italienischen Reden beträgt ca. 6:48 Stunden, die der deutschen Reden ca. 5:58 Stunden (nach Draxler 2023, S. 20). Im Doktorat geht es um Genderunterschiede in politischen Reden in Italien und in Deutschland und das Korpus wurde im Hinblick darauf ausgewählt. Die 44 untersuchten Reden wurden aufgrund unterschiedlicher Anlässe gehalten, bei einigen Reden handelt es sich um Wahlkampfreden, während andere Parlamentsreden sind. Für das Vorhaben dieses Beitrags sowie des Doktorats war es nicht relevant, dass die Reden in derselben Situation stattfanden bzw. zur gleichen Textsorte gehörten, weil sich der Prozess der Transkription nicht ändern würde. Die ausgewählten Reden dienen hier als Beispiel, um verschiedene Web Tools zu präsentieren, Genre und Textsorte stellen in diesem Beitrag keine bedeutende Rolle dar. Die Tools werden vorgestellt, um SprachwissenschaftlerInnen zu unterstützen, die für ihre Untersuchungen Transkripte von (politischen) Reden benötigen.

Die Reden wurden auf der Videoplattform YouTube gesammelt und zunächst durch die kostenlose Software Audacity<sup>23</sup> konvertiert: „the audio from the original videos [gets] converted to wav mono audio files with a 16 bit linear quantization and 16 kHz sample rate“ (Draxler ebd., S. 19). Bei der Namensgebung der Dateien ist in der Regel darauf zu achten, dass die Tools Leerzeichen und Sonderzeichen häufig nicht erkennen. Darüber hinaus ist es sinnvoll, die Dateien mit eindeutig zuzuordnenden Bezeichnungen zu benennen. Dadurch kann man die Reden schnell wiedererkennen und einheitliche Bezeichnungen generieren. Außerdem muss auch beachtet werden, dass die Transkriptions- und Audiofiles der jeweiligen Rede genau denselben Namen haben sollten, sonst werden sie z.B. von WebMAUS nicht erkannt. Im Fall des Korpus wurden die Reden mit Kürzeln aus Namen, Sprache, Geschlecht der SprecherInnen, Partei und Datum der Rede kodiert (ebd., S. 20).

Da YouTube-Videos meist nicht professionell aufgenommen sind, stellen Geräusche, unverständliche Sätze, Unterbrechungen, usw. besondere Herausforderungen dar, was mit einer rein manuellen Transkription kaum zu

<sup>23</sup> <https://www.audacityteam.org/> (04.02.2024).

bewältigen ist. Die hier vorgestellten Web Tools sind daher eine hilfreiche Unterstützung.

Im Fall des vorliegenden Korpus wurden die Reden zunächst automatisch mit Systemen der ASR (Automatic Speech Recognition) transkribiert. Für die italienischen Reden wurde die ASR der Fondazione Bruno Kessler in Trient angewendet, während die deutschen Reden mit dem IBM Watson ASR Service des BAS transkribiert wurden (Draxler 2023, S. 20). Wie Draxler (vgl. ebd.) erwähnt, kann man den IBM Watson ASR Service direkt in OCTRA aufrufen (siehe Abbildung 5, Abschnitt 3.1). Dadurch kann man auf dem Interface von OCTRA die automatischen Transkripte anhand der ASR sehen und sie im Folgenden mit OCTRA korrigieren.

Wissenschaftlich gibt es verschiedene Parameter, um die Effizienz der Transkription messen zu können, so dass man die Leistung mit und ohne ASR vergleichen kann. Draxler (2023, S. 17) benennt in seiner Pilotstudie folgende Faktoren: *raw real-time factor*, *adjusted real-time factor* und *activity factor*.<sup>24</sup> Bei Draxler (ebd.) geht es um technische Aspekte bzw. Ergebnisse und die Studie ist noch vorläufig. Hingegen wird sich dieser Beitrag nicht mit diesen Elementen beschäftigen und daher wird er keine technischen Vergleiche darstellen. Darüber hinaus wurde solch ein Verfahren für Korpora der Politolinguistik bisher nie eingesetzt und diese Methodik wurde das erste Mal im erwähnten Promotionsprojekt verwendet.

Fuß und Karbach (2019, S. 93) verweisen auf die technischen Mängel der Spracherkennungssoftware, die üblicherweise verwendet wird, um Diktate zu erstellen. Diese Programme sind nicht leistungsfähig genug, wenn die Aufnahmequalität schlecht ist oder wenn es Nebengeräusche gibt. Die Autoren (ebd., S. 93-94) legen darüber hinaus dar, dass die Spracherkennungsprogramme es nicht erlauben, Transkriptionszeichen während des Transkriptionsprozess anzubringen. Diese Zeichen können erst am Ende hinzugefügt werden, so dass der Transkriptionsprozess wiederum aufwendig wird. Des Weiteren kann man davon ausgehen, dass die von Fuß und Karbach erwähnten Programme wahrscheinlich keine Möglichkeit bieten, das Audio bzw. Teile des Transkripts zu segmentieren. Außerdem weisen die Autoren in Bezug auf die Spracherkennung mittels künstlicher Intelligenz (ebd., S. 94-95) nur kurz darauf hin, dass noch kein Vergleich mit der manuellen Transkription angestellt wurde. Hier wird auf die künstliche Intelligenz fürs Transkribieren nicht eingegangen, wobei sie sicherlich der ideale nächste Schritt für die Transkription in der (Polito)Linguistik ist.

Dass die ASR eine vielversprechende Ressource sein kann, ist auf jeden Fall mindestens ein Erfahrungswert. Das benannte Forschungsprojekt hat tatsächlich nicht nur den Anspruch, eine politolinguistische Untersuchung, sondern auch eine Lösung für bisherige Schwierigkeiten bei der Transkription

<sup>24</sup> Kursive Markierungen im Original.

zu bieten. Anlässlich der Masterarbeit, die dem Promotionsprojekt vorausgegangen ist, wurde ein Korpus von Wahlkampfreden erstellt, das wegen der Grenzen der manuellen Transkription (in dem Fall mit EXMARaLDA) und des daraus folgenden Zeitaufwands relativ klein war. BAS Web Tools bieten eine alternative Methodik auch für die politische Sprachanalyse, die vielleicht noch nicht mit technischen Details dargestellt werden kann, aber deren Potenzial anfängt, von Studien untersucht zu werden. Da bereits ein Ergebnis aus der vorhergegangenen Erfahrung vorliegt, konnte das Korpus der Doktorarbeit mehr als fünfmal größer sein. Außerdem wurde für die Erstellung des Korpus nur ungefähr ein Drittel der Zeit benötigt. Aus diesen Gründen kann die Verwendung von OCTRA und WebMAUS eine mögliche Alternative darstellen, die sich zwar noch nicht auf bewiesene technische Daten berufen kann, aber zweifelsohne für WissenschaftlerInnen, die Transkripte von langen (politischen) Reden mit Hintergrundgeräuschen benötigen, interessant sein kann.

Für das betrachtete Korpus wurde OCTRA im Online-Modus genutzt und ein Projekt für die Reden erstellt. OCTRA wurde in diesem Fall auch für die Korrekturen von bereits transkribierten Reden verwendet, nachdem die ASR eingesetzt wurde. Im Folgenden sind zwei Beispiele von Transkripten aufgeführt, jeweils mit und ohne Zeitangaben (siehe Abschnitt 3.1), die aus OCTRA exportiert wurden:

<ts="00:03:19.134"/> als der kleine Junge in Frankfurt mutwillig vor den Zug  
gestoßen wurde  
<ts="00:03:22.963"/> traten Sie schweigend Ihren Urlaub an in einem gut  
regierten Land  
<ts="00:03:27.384"/> hätte das Staatsoberhaupt Trauerbeflagung angeordnet  
aber es erfolgte nichts (de\_chru\_m\_afd\_08-11-2019)

als der kleine Junge in Frankfurt mutwillig vor den Zug gestoßen wurde traten  
Sie schweigend Ihren Urlaub an in einem gut regierten Land hätte das  
Staatsoberhaupt Trauerbeflagung angeordnet aber es erfolgte nichts  
(de\_chru\_m\_afd\_08-11-2019)

<ts="00:04:49.650"/> anche come tutti gli altri lavoratori  
<ts="00:04:53.130"/> e poi per carità non faccio il giudice non faccio il  
magistrato non faccio l'avvocato però alcune sentenze ti lasciano senza parole  
(it\_salv\_m\_leg\_01-05-2019)

anche come tutti gli altri lavoratori e poi per carità non faccio il giudice non  
faccio il magistrato non faccio l'avvocato però alcune sentenze ti lasciano senza  
parole (it\_salv\_m\_leg\_01-05-2019)

WebMAUS wurde mit Pipeline without ASR (Abschnitt 3.2) eingesetzt und vier Spuren wurden bearbeitet. Diese können in Abbildungen 13-14 auf Praat und EMU-webApp visualisiert werden. In der Abbildung 13 sind das Signal und die mit WebMAUS bearbeiteten Spuren zu sehen:

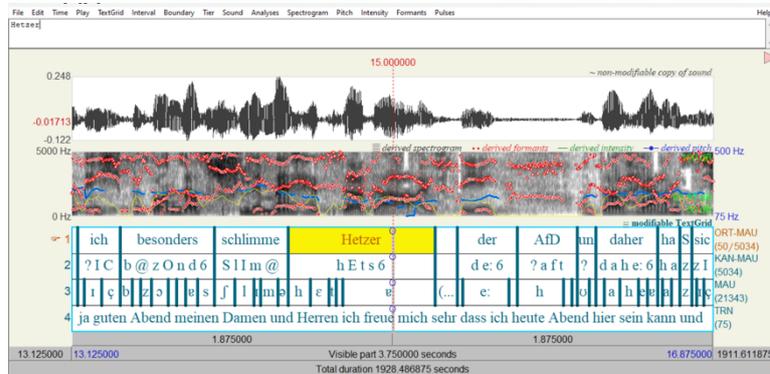


Abbildung 13

Die mit WebMaus alignierten Spuren, die auf Praat geöffnet und analysiert werden können.

Wenn die Transkription mit OCTRA durchgeführt und mit WebMAUS segmentiert bzw. aligniert wird, kann das in einem komplexen Signal mit den diversen Spuren auf EMU-webApp angezeigt werden (Abbildung 14).

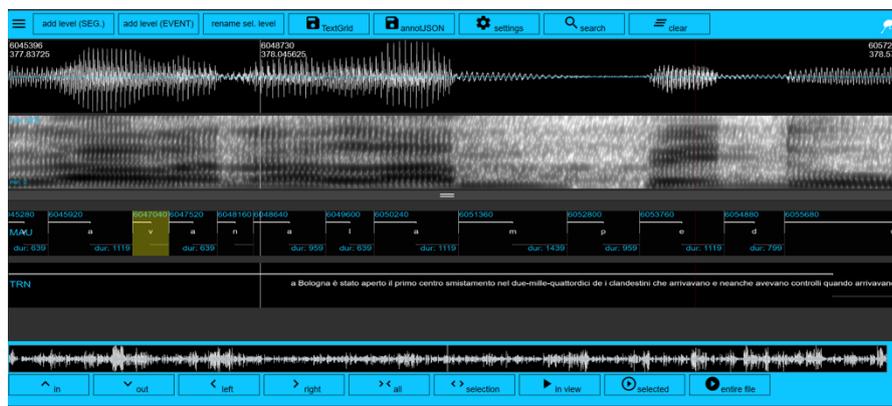


Abbildung 14

Signal und Transkriptionsspuren auf EMU-webApp.

Wenn EMU-webApp direkt im Browser verwendet und nicht über emuR aufgerufen wird, werden nicht alle vier Spuren angezeigt. Dies ist zweifelsohne ein Aspekt, der durch weitere Studien und Entwicklungen verbessert werden kann. Wie bereits erwähnt, wird R in sprachwissenschaftlichen Disziplinen selten benutzt und es wäre wünschenswert, EMU-webApp auch direkt im Browser nutzen zu können, ohne dass die Visualisierung der mit WebMAUS erstellten Spuren beeinträchtigt wird.

## 5. Fazit

Es gibt zur Zeit zahlreiche Möglichkeiten, die orthographische Transkription mittels verschiedener Software und Tools durchzuführen, und die in diesem Beitrag näher beleuchteten Anwendungen bieten eine zwar noch wenig bekannte, aber umso spannendere und nützliche webbasierte Alternative zu bisher genutzten Softwareprogrammen. Im Beitrag wurden die Vor- und Nachteile, sowie die Unterschiede zu herkömmlichen Methoden erörtert, um aufzuzeigen, inwiefern sie auch in anderen Disziplinen und Bereichen der Linguistik, wie der (Polito)Linguistik, Anwendung finden könnten.

Üblicherweise liegt der Analysefokus in der Linguistik entweder auf sprachwissenschaftlichen Eigenschaften oder auf dem Signal. Es gibt meist keinen Kompromiss bzw. Verknüpfungen sprachwissenschaftlicher Interessensschwerpunkte. Nichtsdestotrotz sind diese zwei Ebenen vor allem in der politischen Sprache eng miteinander verbunden: Wie man sich äußert, trägt zur Wirksamkeit der Inhalte der Reden bei und obwohl die politischen Programme sowie oft auch einige Reden schriftlich zur Verfügung stehen, besteht die politische Kommunikation vor allem aus der gesprochenen Sprache. Aus diesem Grund wird eine Analyse, die diese beiden Ebenen verknüpft, zahlreiche Aspekte vertiefen.

Die präsentierten Tools könnten an die verschiedenen Forschungszwecke in der (Polito)Linguistik angepasst werden und die Aufgabe der Transkription und der Alignierung bzw. Segmentierung schneller und unkomplizierter machen.

**Bionote:** Marcella Palladino is a PhD student at the Università degli Studi di Modena e Reggio Emilia (UNIMORE). Her research focus is on German Linguistics, especially politolinguistics, rhetoric, critical discourse analysis and prosody. During her PhD, she has been Visiting PhD Student at the Institut für Phonetik und Sprachverarbeitung (IPS), LMU in Munich. Her doctoral project deals with the analysis of voice and prosodic characteristics of male and female politicians in Italy and Germany.

**Author's address:** [marcella.palladino@unimore.it](mailto:marcella.palladino@unimore.it)

**Acknowledgements:** I would like to thank PD Dr. Christoph Draxler and the Institut für Phonetik und Sprachverarbeitung (IPS) at the Ludwig-Maximilians-Universität München for their precious support to the current research, and Dr. Daniele Falavigna and his team at the Fondazione Bruno Kessler in Trento for the ASR system used for the Italian speeches.

## Literatur

- Brambilla M.M. 2007, *Il discorso politico nei paesi di lingua tedesca: metodi e modelli di analisi linguistica*, Aracne, Roma.
- Burkhardt A., 1996 *Politolinguistik. Versuch einer Ortsbestimmung*, in Diekmannshenke, Hajo/Klein, Josef (Hrsg.), *Sprachstrategien und Dialogblockaden. Linguistische und politikwissenschaftliche Studien zur politischen Kommunikation*, Berlin/New York: de Gruyter, pp. 75-100.
- Draxler C. 2023, *Analysis of transcriptions using Octra – a pilot study*, in Draxler C. (Hrsg.), *Elektronische Sprachsignalverarbeitung 2023*, 105, TUDpress, Dresden, pp. 17-23.
- Fuß. S. and Karbach U. 2019, *Grundlagen der Transkription*, 2. Aufl., Verlag Barbara Budrich, Opladen & Toronto.
- Heilmann C.M. 2003, *Von der Mündlichkeit über die Schriftlichkeit zu veränderter Mündlichkeit. Vom Sinn des Transkribierens*, in Anders L.C., Hirschfeld U. (Hrsg.), *Sprechsprachliche Kommunikation. Probleme, Konflikte, Störungen*, in „Hallesche Schriften zur Sprechwissenschaft und Phonetik“ 12, Peter Lang, Frankfurt/M., pp. 141-151.
- Kirk J.M. and Andersen G. 2016, *Compilation, transcription, markup and annotation of spoken corpora*, in “International Journal of Corpus Linguistics” 21, pp. 291-298. <https://doi.org/10.1075/ijcl.21.3.001int> (15.10.2023).
- Kisler T., Reichel U.D., Schiel F., Draxler C., Jackl B. and Pörner N. 2016, *BAS Speech Science Web Services – an Update on Current Developments*. [https://www.researchgate.net/publication/303750698\\_BAS\\_Speech\\_Science\\_Web\\_Services\\_-\\_an\\_Update\\_on\\_Current\\_Developments](https://www.researchgate.net/publication/303750698_BAS_Speech_Science_Web_Services_-_an_Update_on_Current_Developments) (25.04.2023).
- Kisler T., Reichel U.D. and Schiel F. 2017, *Multilingual processing of speech via web services*, in “Computer Speech & Language” 45, pp. 326-347.
- Niehr T. 2014, *Einführung in die Politolinguistik. Gegenstände und Methoden*, Vandenhoeck & Ruprecht, Göttingen.
- Poerner N. and Winkelmann R. 2018, *Interfacing the BAS Speech Science Web Services and the EMU Speech Database Management System*, in “P&P13 – Proceedings of the Conference on Phonetics and Phonology in German-speaking countries”, Berlin, pp. 141-144.
- Pömp J. and Draxler C. 2017, *OCTRA – a Configurable Browser-Based Editor for Orthographic Transcription*, in “P&P13 – Proceedings of the Conference on Phonetics and Phonology in German-speaking countries”, Berlin, pp. 145-148.
- Schmidt T. 2016, *Good practices in the compilation of FOLK, the Research and Teaching Corpus of Spoken German*, in “International Journal of Corpus Linguistics” 21, pp. 291-298. <https://doi.org/10.1075/ijcl.21.3.001int> (15.10.2023).
- Selting M., Auer P., Barth-Weingarten D., Bergmann J., Bergmann P., Birkner K., Couper-Kuhlen E., Deppermann A., Gilles P., Günthner S., Hartung M., Kern F., Mertzlufft C., Meyer C., Morek M., Oberzaucher F., Peters J., Quasthoff U., Schütte W., Stukenbrock A. and Uhmann S. 2009, *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)*, in „Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion“ 10, pp. 353-402. <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf> (15.10.2023).

## Tools und Software

- Audacity (Version 3.1.3) [Computer software] 1999-2016, Audacity Team. <https://www.audacityteam.org/> (06.02.2024).
- Boersma P., Weenink D. 2022, *Praat: doing phonetics by computer version 6.2.06* [Computer program]. <https://www.praat.org/>. (05.12.2023).
- Brugman H., Russel A. 2004, *Annotating Multimedia/ Multi-modal resources with ELAN*, in “Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation”.
- ELAN (Version 6.7) [Computer software] 2023, Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. <https://archive.mpi.nl/tla/elan> (05.12.2023).
- Jochim M., Winkelmann R., Jaensch K., Cassidy S., Harrington J. 2023. *emuR: Main Package of the EMU Speech Database Management System. R package version 2.4.2.* <https://CRAN.R-project.org/package=emuR> (09.02.2024).
- R Core Team 2023. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org> (09.02.2024).
- Schiel F. 1999, *Automatic Phonetic Transcription of Non-Prompted Speech*, in “Proceedings of the ICPhS 1999”, San Francisco, pp. 607-610.
- Schiel F. 2015, *A Statistical Model for Predicting Pronunciation*, in “Proceedings of the ICPhS 2015”, Glasgow, UK, paper 195.
- Schmidt T. 2020, *Exmaralda 1.6* [Computer software]. Universität Hamburg, IDS Mannheim (2011). <https://exmaralda.org/de/ueber-exmaralda/> (05.12.2023).
- Schmidt T., Schütte W. 2010, *FOLKER: An Annotation Tool for Efficient Transcription of Natural, Multi-party Interaction*, in “Proceedings of LREC 2010”. [https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/2232/file/Schmidt\\_Schuette\\_FOLKER\\_2010\\_Paper.pdf](https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/2232/file/Schmidt_Schuette_FOLKER_2010_Paper.pdf) (05.12.2023).
- Winkelmann R., Harrington J. and Jänsch K. 2017, *EMU-SDMS: Advanced speech database management and analysis in R*, in “Computer Speech & Language” 45, pp. 392-410.