



Oral-History.Digital

## **Automatisches Alignment von Oral History-Interviews.**

Anleitung für die Nutzung der BAS Web Services

**Herdis Kley, Cord Pagenstecher, Florian Schiel unter Mitarbeit von Tobias Kilgus, Peter Kompiel und Philipp Linß**

Erarbeitet im Rahmen des Projekts „Oral-History.Digital“, gefördert von der Deutschen Forschungsgemeinschaft (DFG), vom Bayerischen Archiv für Sprachsignale an der Ludwig-Maximilians-Universität München und dem Bereich Digitale Interview-Sammlungen an der Universitätsbibliothek der Freien Universität Berlin

Version 0.5 | 15.12.2022

[www.oral-history.digital](http://www.oral-history.digital)

# Automatisches Alignment von Oral History-Interviews.

## Anleitung für die Nutzung der BAS Web Services

Ton- und Videoaufnahmen sind die Kernelemente einer Interviewsammlung, aber viele Funktionalitäten wie Suche, Navigation und Annotation sind immer noch textbasiert. Oftmals liegen bereits Transkripte zu den audiovisuellen Aufnahmen vor, die aber erst noch über Timecodes miteinander verbunden werden müssen. Erst diese Segmentierung (bzw. Alignment) erlaubt eine synchrone Untertiteldarstellung und eine sekundengenaue Volltextsuche oder Annotation. Dabei wird nach bestimmten Zeitabständen (z.B. eine Minute), nach einer gewissen Wort- oder Zeichen-Anzahl oder nach Sinneinheiten (z.B. nach jedem Satz) ein Timecode eingefügt.

Um die Audio- und Videodateien durchsuchbar zu machen und die Transkripte (oder Übersetzungen) als Untertitel anzuzeigen, müssen die Transkripte mit Timecodes segmentiert und mit den Mediendateien gekoppelt werden. Auch für dieses sogenannte Alignment gibt es Software-Unterstützung.

Im Projekt *Oral-History.Digital* werden Software-Werkzeuge für die Spracherkennung (Automatic Speech Recognition, ASR) und für die automatische Segmentierung (Alignment) vom Bayerischen Archiv für Sprachsignale an der Ludwig-Maximilians-Universität München geprüft, angepasst und weiterentwickelt. Der Bereich Digitale Interview-Sammlungen an der Universitätsbibliothek der Freien Universität Berlin erarbeitet Handreichungen zur Nutzung dieser Dienste für Oral History-Interviews und prüft die Einrichtung von Schnittstellen aus der Erschließungsplattform *oral-history.digital*. Bitte beachten Sie, dass es aufgrund der Weiterentwicklung beim BAS und den externen Spracherkennern regelmäßig zu Änderungen der Nutzungsbedingungen und Anforderungen kommen kann, so dass diese Anleitung nur den momentanen Stand berücksichtigt und ohne Gewähr ist. Bitte informieren Sie sich vor der Anwendung der BAS Web Services über die aktuellen Nutzungsbedingungen auf deren Webseite: <https://clarin.phonetik.uni-muenchen.de/BASWebServices/help/termsOfUsage>.

Für das Automatische Alignment (Segmentierung) nutzen Sie direkt die BAS Webservices auf <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface>. Die folgende Anleitung unterstützt Sie dabei, für ein bereits vorhandenes Transkript zu einer Audio- oder Videodatei eines Oral History-Interviews Timecodes zu erzeugen. Die so erzeugte Untertiteldatei im vtt-Format können Sie in allen gängigen Medienplayern direkt anzeigen lassen, in die Erschließungsplattform *oral-history.digital* importieren oder in einem geeigneten Transkriptionsprogramm, wie zum Beispiel InqScribe, nachbearbeiten.

## Vorbereitung vorhandener Transkripte

Die BAS Webservices unterstützen eine Vielzahl an Sprachen für die unterschiedlichen Services. Bitte prüfen Sie im Dropdown-Menü der Sprachauswahl bei den jeweiligen Services, ob die in Ihren Interviews verwendete Sprache verarbeitet werden kann.

Bereiten Sie das vorhandene Transkript sorgfältig vor, um möglichst passende Timecodes zu erhalten. Das automatische Alignment setzt Timecodes nach jedem finalen Satzzeichen (Punkt, Doppelpunkt, Auslassungspunkte, Ausrufezeichen, Fragezeichen), daher sollten Abkürzungen, die Punkte enthalten, aufgelöst und Datumsangaben mit Punkten oder Zeitangaben mit Doppelpunkten maskiert werden. Sie können die Bearbeitung problemlos in Word vornehmen (Suchen-Ersetzen-Funktion), da u.a. folgende Formate erlaubt sind: csv, doc, docx, odt, par, pdf, rtf, txt.

**Achtung:** Es hat sich herausgestellt, dass das rtf-Format manchmal nicht korrekte Formatierungsbefehle verwendet und dann Sprecherkürzel nicht erkannt werden. Es empfiehlt sich daher, rtf-Dateien für das Alignment beim BAS vorher in eines der anderen Textformate abzuspeichern.

1. Fügen Sie bei Sprecherwechsel das <Sprecherkürzel> in spitzen Klammern ein. Bitte beachten Sie, dass das Sprecherkürzel aus **mindestens** 2 Buchstaben bestehen muss.  
**Beispiel:** <INT>, <AS>, <UL>
2. Punkt, Ausrufezeichen, Fragezeichen, Doppelpunkt und Auslassungspunkte werden als finale Satzzeichen gewertet und erzwingen eine Segmentierung (d.h. es wird ein neuer Timecode gesetzt). Ergänzen Sie daher ggf. fehlende finale Satzzeichen vor einem Sprecherwechsel.  
**Achtung:** Bei unterbrochenen Sätzen den Satz nicht ohne Satzzeichen beenden, sondern mit Auslassungspunkten (...) enden.  
**Beispiel:** <INT> Ach so? <AS> Ja, leider... <INT> Das ist schlimm!  
**Hinweis:** Durch Suchen und Ersetzen mit Absatzmarke (^p) können schnell fehlende Satzzeichen vor Sprecherwechseln ergänzt werden:
  1. Nach ^p suchen und mit .^p ersetzen. (In Word befindet sich die Absatzmarke im „Suchen und Ersetzen-Fenster“ unter „Erweitern“ bei „Sonderformat“). Daraufhin wird vor jedem Absatz (Sprecherwechsel) ein weiterer Punkt gesetzt.
  2. Dadurch doppelt gesetzte Punkte können dann wieder gelöscht werden: Suchen nach ..^p und ersetzen mit .^p und ggf. für weitere finale Satzzeichen wiederholen, zum Beispiel ?.^p ersetzen mit ?^p oder !.^p ersetzen mit !^p
3. Maskieren Sie Datumsangaben, die Punkte enthalten, mit spitzen Klammern.  
**Beispiel:** <19.02.1956>, <13.> März
4. Löschen Sie vorhandene Timecodes oder maskieren Sie diese mit spitzen Klammern.  
**Beispiel:** <00:04:34> Bitte beachten Sie, dass die Timecodes im Satz stehen müssen, damit kein Umbruch vor dem maskierten Timecode gesetzt wird: **nicht** „...und machte dann zu. <00:04:34>“ **sondern** „...und machte dann zu <00:04:34> .“  
**Achtung:** Bei maskierten Timecodes am Satzende, muss vor und nach der spitzen Klammer ein Leerzeichen eingefügt werden.
5. Maskieren Sie alle weiteren Anmerkungen (d.h. alle Informationen, die nicht im Interview gesprochen werden) mit spitzen Klammern.  
**Beispiel:** <Lachen>, <Telefonklingeln>

Diese maskierten Anmerkungen dürfen nicht an Wörtern 'kleben', d.h. zwischen einem Wort und dem '<...>' muss ein Leerzeichen stehen.

**Beispiel:** Ich bin in München **geboren** <lacht nervös>.  
Ich warte im **Unter** <räuspert> **geschoß** auf dich.

6. Lösen Sie Abkürzungen so auf, dass keine Punkte mehr verwendet werden.

**Beispiel:** »usw.« zu »und so weiter« / »d.h.« zu »das heißt«

Übersicht der häufigsten Abkürzungen:

Abh.	Abhandlung	i. e. S.	im engeren Sinne
Abk.	Abkürzung	i. w. S.	im weiteren Sinne
allg.	allgemein	jmd.	jemand
bes.	Besonders	o. Ä.	oder Ähnliche(s)
bez.	bezahlt, bezüglich	u.	und
bsw.	beispielsweise	u. a.	und anderes, unter anderem
bzw.	beziehungsweise	u. Ä.	und Ähnliche(s)
ca.	circa	u. dgl.	und dergleichen
d. h.	das heißt	ugs.	umgangssprachlich
eigtl.	eigentlich	urspr.	ursprünglich
geb.	geboren	usw.	und so weiter
gegr.	gegründet	z. B.	zum Beispiel
ggf.	gegebenenfalls	zz./zzt.	zurzeit
Ggs.	Gegensatz		

7. Löschen Sie ggf. Kopf- und Fußzeilen.  
8. Löschen Sie ggf. Zwischenüberschriften und alle Textpassagen, die nicht gesprochen wurden.  
9. Maskieren Sie ggf. Auslassungspunkte, die in der Mitte eines Satzes stehen, mit spitzen Klammern.  
10. Benennen Sie die Transkript-Datei identisch mit der Medien-Datei, z.B. »interview1.docx« + »interview1.wav«

Beispiel für ein <b>schlecht</b> vorbereitetes Transkript	<INT> Dann fangen wir mal an. Wie ist Ihr Name?(00:04.34) <sup>1</sup> <TAR> <sup>2</sup> Ich heiÙe Michaela Mustermann und bin in München geboren. <lacht nervös> <INT> Und<P>Papier raschelt laut>wann? <sup>3</sup>
Beispiel für ein <b>optimal</b> vorbereitetes Transkript:	<INT> Dann fangen wir mal an. Wie ist Ihr Name <00:04.34>? <TAR> Ich heiÙe Michaela Mustermann und bin in München geboren <lacht nervös>. <INT> Und <P>Papier raschelt laut> wann?
Korrektur	<sup>1</sup> Setzen Sie den vorhandenen Timecode vor das Fragezeichen und ersetzen Sie die Klammern durch spitze Klammern. <sup>2</sup> Vor dem Sprecherwechsel TAR steht kein terminierendes Satzzeichen, sondern eine Klammer. <sup>3</sup> Fügen Sie Leerzeichen vor und nach den spitzen Klammer hinzu. Sprecher-ID INT hat kein Satzzeichen davor, sondern eine Klammer.

## Vorbereitung der Medien-Dateien

Um ein schnelles Hochladen zu ermöglichen, empfiehlt es sich bei Videodateien die Tonspur herauszuspielen und als WAV-Datei mit mindestens 16kHz und in Mono abzuspeichern. Wenn das nicht möglich ist, kann man auch das Video selber an den BAS-Server schicken (dauert länger und wird dann auf dem Server automatisch transkodiert). Folgende Formate sind erlaubt: aiff, au, avi, flac, flv, mpg, mp3, mpeg, mp4, nis, nist, ogg, snd, sph, wav.

Benennen Sie die Medien-Datei identisch mit der Transkript-Datei: »interview1. docx« und »interview1.wav«

Bitte beachten Sie, dass eine maximale Dateigröße von 2 GB (Gesamtmenge für alle Dateien einer Verarbeitung) nicht überschritten werden darf.

Ihre Daten werden im Zuge der Bearbeitung durch die BAS Webservices auf deren Servern (an der Ludwig-Maximilians-Universität München) gespeichert und bearbeitet und automatisch nach 24 Stunden wieder gelöscht. Eine Weitergabe an Dritte erfolgt nicht.

## Alignment vorhandener Transkripte (Pipeline w/out ASR)

**Achtung:** Transkript und Medien-Datei müssen nach den oben genannten Maßgaben formatiert und identisch benannt sein, z.B. »interview1.docx« und »interview1.wav«

1. Starten Sie Google Chrome und gehen Sie auf die Webseite **Pipeline w/out ASR:**  
<https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/Pipeline>
2. Laden Sie die Medien-Datei **und** die Transkript-Datei hoch.

**Bemerkung:** Das Hochladen kann je nach Datei-Größe einige Minuten bis zu einer Stunde dauern.

**Hinweis:** Falls sie gleichzeitig mehrere Dateien verarbeiten möchten, können Sie an dieser Stelle mehrere Dateien paarweise hochladen. Bitte öffnen Sie dafür kein weiteres Browserfenster, um parallel mehrere Dateien an die BAS Spracherkennung zu schicken.

3. Geben Sie unter **Service Options** folgende Werte ein:
  - a. Pipeline name: **G2P-CHUNKER-MAUS**
  - b. Language: **[Wählen Sie die Sprache des Interviews aus]**
  - c. Output format: **BAS Partitur Format (bpf)**

The screenshot displays the 'BAS Web Services' interface for the 'Pipeline without ASR' service. The main content area is divided into several sections:

- Files:** A list of three files successfully uploaded, each consisting of a .docx file and a corresponding .wav audio file (e.g., interview\_1.docx <-> interview\_1.wav).
- Service options:** A form with three dropdown menus: 'Pipeline name (required)' set to 'G2P-CHUNKER-MAUS', 'Language (required)' set to 'German (DE)', and 'Output format (required)' set to 'BAS Partitur Format (bpf)'. There is also a 'Keep everything' checkbox.
- Expert Options:** A section with a 'click to show' link and a 'Run' button.
- Service manual:** A section on the right with a 'hide >' button, containing instructions on how to use the service and a list of supported file formats.

At the bottom, there is a 'Run' button and a 'Server Load Indicator' showing a green bar.

4. Klicken Sie auf Expert Options.

**Bemerkung:** Es erscheint eine lange Liste von Optionen, in der Sie Folgendes ergänzen:

- a. Geben Sie bei **User email notification** Ihre E-Mail-Adresse ein.

**Bemerkung:** Die Arbeit des Web Service kann mehrere Stunden dauern. Sie erhalten eine E-Mail »Your Pipeline Job has finished« mit einem Download-Link zugeschickt (nur für 24h gültig!), wenn der Web Service Ihren Auftrag bearbeitet hat. Sie können jetzt den Browser schließen. Wenn Sie Ihre E-Mail-Adresse eingegeben haben, können Sie den Browser nach dem Start der Verarbeitung beenden.

## Pipeline without ASR

Expert Options (click to hide):

Input tier name (optional)(G2P,CHUNKPREP)	unknown	?
Input Encoding (CHUNKER,MAUS)	X SAMPA	?
Keep annotation (G2P,CHUNKPREP)	yes	?
Output Encoding	X-SAMPA (ASCII)	?
User email notification	m.muster@uni-muster.de	?
Audio input pre-processing (AUDIOENHANCE)	true	?

b. Geben Sie bei der Option **Annotation marker brackets** alle Klammern ein, die Sie im Transkript verwenden, um **nicht gesprochene Informationen** zu markieren, z.B. '<>{}[]()'.  
**Bemerkung:** Die Eingabe von '{}[]()' ist nur notwendig, wenn im Transkript noch andere Klammern als die üblichen '<>' vorkommen, um **nicht gesprochenen** Text zu maskieren. Wenn Sie nur '<>' im Text verwenden, müssen Sie nichts eingeben (default).

c. Geben Sie bei der Option **Comment character** alle Zeichen ein, die Sie im Transkript verwenden, um Kommentarzeilen zu markieren, z.B. ';' (default ist '#').  
**Bemerkung:** Kommentarzeilen sind Informationen zum Interview, die nicht gesprochen wurden und meistens mit einem eindeutigen 'Kommentarzeichen' am Beginn der Zeile markiert werden, z.B.:

- # Interview No 21.
- # Datum: 2022-03-01.
- # Interviewer: Max Mustermann.

**Achtung:** Falls Sie # im Transkript benutzen, und dies **nicht** eine Kommentarzeile einleitet, ändern Sie bitte hier den default-Wert in NONE.

## Pipeline without ASR

Output frame rate (MAUS)	10msec	?
Add Viterbi likelihoods (MAUS)	false	?
Text input pre-processing (TEXTENHANCE)	true	?
Utterance level modelling (TEXTENHANCE)	false	?
Annotation marker brackets (TEXTENHANCE)	<>{}[]()	?
Comment character (TEXTENHANCE)	#	?
White space replacement (TEXTENHANCE)	_	?

**Run**

I have read and accepted the [terms of usage](#) for this service, including the policy of monitoring access to the services (paragraph 5). I hereby confirm that I am a member of an academic institution or that I have obtained a BAS user license for this service. In case of a publication of my results I will use a proper citation to this service.



<div style="background-color: #0070C0; color: white; padding: 2px; text-align: center; margin-bottom: 5px;">Copy to clipboard</div> <div style="background-color: #0070C0; color: white; padding: 2px; text-align: center; margin-bottom: 5px;">Reset color</div> <div style="background-color: #0070C0; color: white; padding: 2px; text-align: center; margin-bottom: 5px;">Clear messages</div> <input checked="" type="checkbox"/> Errors <input checked="" type="checkbox"/> Warnings <input checked="" type="checkbox"/> Success	Background color: <input type="checkbox"/> Error <input type="checkbox"/> Warning <input type="checkbox"/> Success <input type="checkbox"/> No messages  + (19:04:12.069) Success: Upload was successful + (18:58:43.541) WARNING: Please consider "user email notification" for large files
--	---

- Akzeptieren Sie die **Nutzungsbedingungen** und klicken Sie den Button **Run Web Service** an.

**Achtung:** Ihre Daten werden im Zuge der Bearbeitung durch die BAS Webservices auf deren Servern (an der Ludwig-Maximilians-Universität München) gespeichert und bearbeitet und automatisch nach 24 Stunden wieder gelöscht.

- Rechts-Klick auf den PAR-Ergebnis-Link (.par ist die Dateierweiterung für das BAS Partitur Format) lädt die PAR-Datei auf Ihren PC herunter.

\*MacOS- und Googlemail-Nutzer\*innen beachten bitte den Hinweis am Ende der Anleitung.

**Achtung:** Wenn Sie eine E-Mail ohne Ergebnislink, aber mit einer „Error“-Meldung erhalten, leiten Sie diese E-Mail bitte an die Entwickler des BAS weiter (web-services@bas.uni-muenchen.de).

- Bearbeiten Sie die PAR-Datei nach der Anleitung **Untertitel/Transkript erstellen**.

**Bemerkung:** Der Name der PAR-Datei entspricht standardmäßig den Input-Dateien: »interview1.par«

## Mögliche Warnungen/Fehlermeldungen der Pipeline:

Es kann vorkommen, dass Sie eine PAR-Datei erhalten, in der Subtitles nicht richtig aligniert werden konnten. In solchen Fällen erhalten Sie mit dem Ergebnis-Link eine Warnung.

### Chunker-Probleme (notProcessedChunk)

- **WARNING:** maus.trn : no results for chunk number 1 (begins at 0sec and lasts for 1931secs) - inserting equally timed dummy (word) labels <notProcessedChunk> in MAU tier; if this MAUS call is part of a pipeline that end in service SUBTITLE, all subtitles in this time range will be faulty!
- **Lösungsansatz:** Lassen Sie die Pipeline ohne den Chunker laufen, indem Sie alle Parameter wie oben beschrieben einstellen und nur bei **3a Pipeline name: G2P -MAUS** auswählen.

**Achtung:** Bitte beachten Sie, dass die Pipeline G2P\_M AUS nur bei einem Input mit weniger als 10.000 Wörtern funktioniert und deutlich länger dauert.

### Chunker-Versagen (Abbruch der Pipeline)

- **ERROR:** Result TRN tier contains a turn with a length beyond 10000 words. This means that MAUS (or other forced aligners) might still have trouble segmenting the material in real time. Consider lowering the minanchorlength parameter or running the chunker with the force option set to true
- **Hinweis:** Der Chunker-Dienst versucht, transkribierte Signale zusammen mit dem Transkript in Teile aufzuteilen, die kürzer als 10.000 Wörter sind, so dass die MAUS sie verarbeiten kann. Wenn diese Aufteilung fehlschlägt, d.h. es gibt mindestens einen Teil nach der Aufteilung, der länger als 10000 Wörter ist, stoppt die Pipeline, weil dieser so genannte Chunk nicht weiterverarbeitet werden kann.  
Leider gibt es keine einfache Abhilfe für diesen Fehler. Der Grund, warum der Chunker fehlschlägt, ist oft (aber nicht immer) ein Missverhältnis zwischen Signal und Transkript oder eine sehr schlechte Signalqualität (viel Hintergrundrauschen oder Übersprechen).
- **Lösungsansatz:** Manchmal hilft es, die CHUNKER-Option 'Use MAUS word model (CHUNKER)' oder die Option Forced chunking (CHUNKER) auf „true“ zu setzen.

**Achtung:** Bitte beachten Sie, dass die Verarbeitung von großen ‚Chunks‘ oftmals dazu führt, dass die Timecodes innerhalb dieser Chunks auseinanderlaufen.

## Untertitel/Transkript erstellen (VTT-Output)

**Achtung:** Es müssen alle Schritte der Anleitung Alignment der vorhandenen Transkripte (Pipeline w/out ASR) erfolgt sein.

1. Starten Sie Google Chrome und gehen Sie auf die Webseite **Subtitle**:  
<https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/Subtitle>
2. Laden Sie die PAR-Datei(en) und das/die vorhandene(n) Transkript(e) hoch.  
**Beispiel:** »interview1.par« **und** »interview1.docx«
3. Geben Sie unter **Service Options** folgende Werte ein:
  - a. Maximum subtitle length: **0**  
**Bemerkung:** Maximum subtitle length: 0 bedeutet, dass neue Segmente erst nach einem abschließenden Satzzeichen (Punkt, Doppelpunkt, Ausrufezeichen, Fragezeichen, Auslassungspunkte) gebildet werden. Wenn Sie zusätzliche Timecodes nach einer maximalen Anzahl von Wörtern benötigen, ändern Sie einfach den Wert (1, 10, 30...).
  - b. Output format: **vtt**  
**Bemerkung:** Sie können zwischen folgenden Ausgabeformaten wählen: srt, sub, vtt, bpf+trn

Show service sidebar > **BAS Web Services**  
Version 3.11 • History of changes

### Subtitle

**Files**

Files successfully uploaded:

- 1. interview\_1.docx <=> interview\_1.par
- 2. interview\_2.docx <=> interview\_2.par
- 3. interview\_3.docx <=> interview\_3.par

Delete all

**Service options**

Maximum subtitle length: 0

Subtitle split marker: punct

Output format: vtt

**Run**

I have read and accepted the [terms of usage](#) for this service, including the policy of monitoring access to the services (paragraph 5). I hereby confirm that I am a member of an academic institution or that I have obtained a BAS user license for this service. In case of a publication of my results I will use a proper citation to this service.

Run Web Service

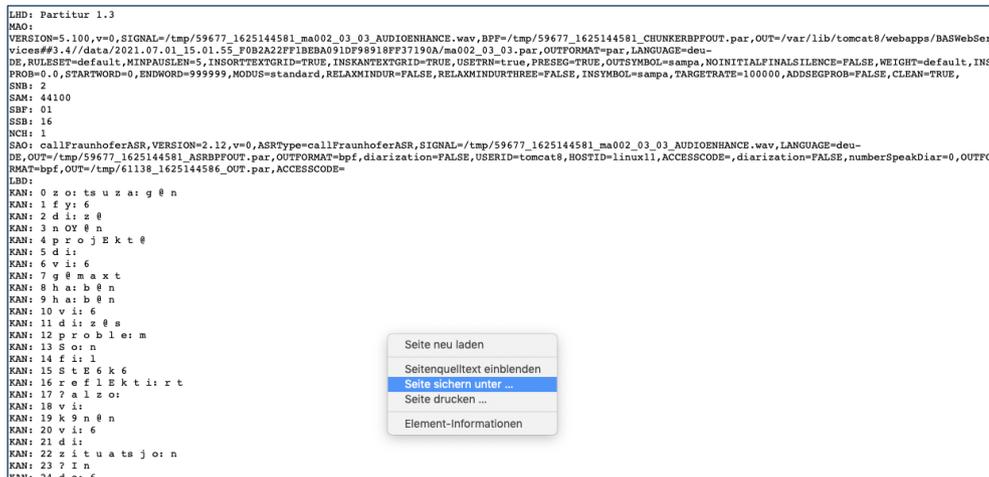
4. Akzeptieren Sie die **Nutzungsbedingungen** und klicken Sie den Button **Run Web Service** an.  
**Bemerkung:** Der Subtitle Service benötigt i.d.R. nur einige Sekunden/Minuten.
5. Downloaden und speichern Sie das Ergebnis durch Rechts-Klick auf den VTT-Ergebnis-Link.  
(\*MacOS- und Gmail-Nutzer\*innen beachten bitte den Hinweis am Ende der Anleitung.)

Die so erzeugte Untertiteldatei im vtt-Format können Sie in allen gängigen Medienplayern direkt anzeigen lassen oder in die Erschließungsplattform von Oral-History.Digital importieren.

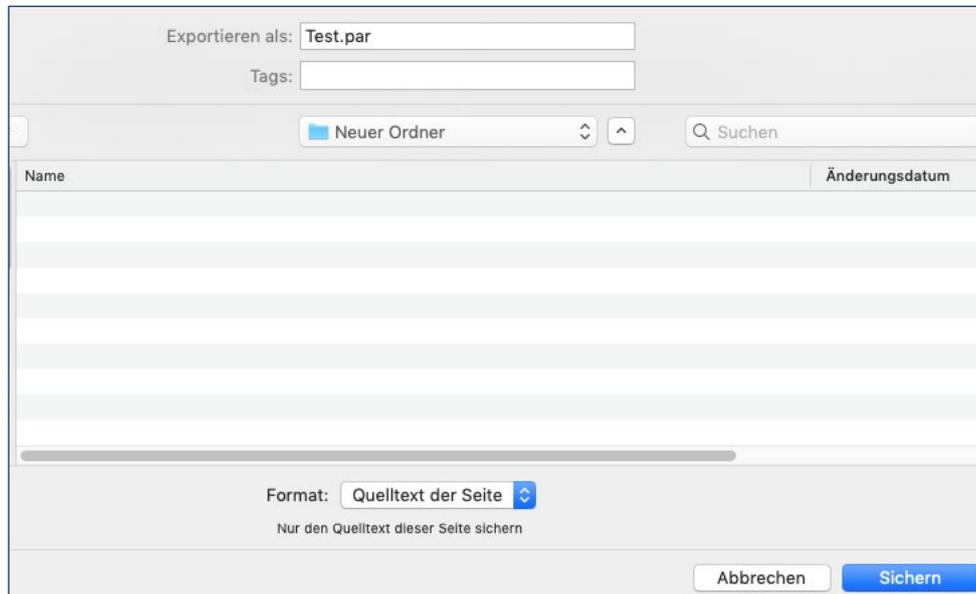
### \*Hinweis für MacOS- und Gmail-Nutzer\*innen:

In Apple Mail lässt sich die Datei nicht über einen Rechts-Klick herunterladen. Daher muss man den Quelltext der verlinkten Webseite über den Safari-Browser herunterzuladen.

Dafür wählt man auf der Webseite Rechtsklick -> **Seite sichern unter**



Als Format "Quelltext der Seite" wählen und als Dateiendung **.par** (bzw. **.vtt**) eintippen.



**Wichtig:** bei der folgenden Meldung "nicht anfügen" auswählen.

