Oral History & Technology

# Handleiding WhisperCorrector

v.a. 3.0.5 ev, 9 april. 2026

*Met ondersteuning van het PDI-SSH Project Oral History - Stories at the Museum Around Artworks (OH-SMArt)*

## WhisperCorrector

WhisperCorrector-software is bedoeld voor het lokaal corrigeren van resultaten van Automatische Spraakherkenning (Automatic Speech Recognition - ASR) in *.srt en/of *.json formaat, zoals die van Whisper, vandaar de naam WhisperCorrector. Je kunt er veilig privacygevoelig materiaal op je eigen computer bewerken. Je kunt spreker namen toevoegen, tekst corrigeren, woorden in één keer vervangen, leestekens toevoegen, en het resultaat in 4 verschillende formaten exporteren:

1. *.srt, een ondertitelingsformaat, evt later om te zetten in *.vtt;
2. *.html, een doorzoekbare audio-tekst interface;
3. *.txt, een rekenkundig goed leesbaar tekstbestand;
4. *.word.txt, met vooringestelde opmaak, makkelijk te copy-pasten in Word voor leesbaar transcript met sprekers voor de tekst, door linker kantlijn 3 cm naar rechts schuiven.

WhisperCorrector is open source software, gratis te downloaden van https://sprekendegeschiedenis.nl, onder Aan de slag / Tools, voor zowel Windows (zip-file) als Apple (dmg-file). Afhankelijk van de Privacy Settings op je computer, zal je eerst toestemming moeten geven om het bestand te openen (negeer de pop-up suggestie het programma te verwijderen), bijvoorbeeld onder Systeeminstellingen/Privacy en beveiliging, kies: Open toch.

## 1. Importeren en instellen

Om te beginnen zul je éénmalig het transcriptieresultaat moeten importeren uit de map met resultaten. Zorg dat de opname, het AV-bestand (audio of video) in dezelfde map blijft staan en dezelfde bestandsnaam heeft, net als alle files in deze map zodat ze met elkaar corresponderen.

Open WhisperCorrector en ga éénmalig naar Settings links in het menu, en klik aan: Expert Mode (Fig. 1).

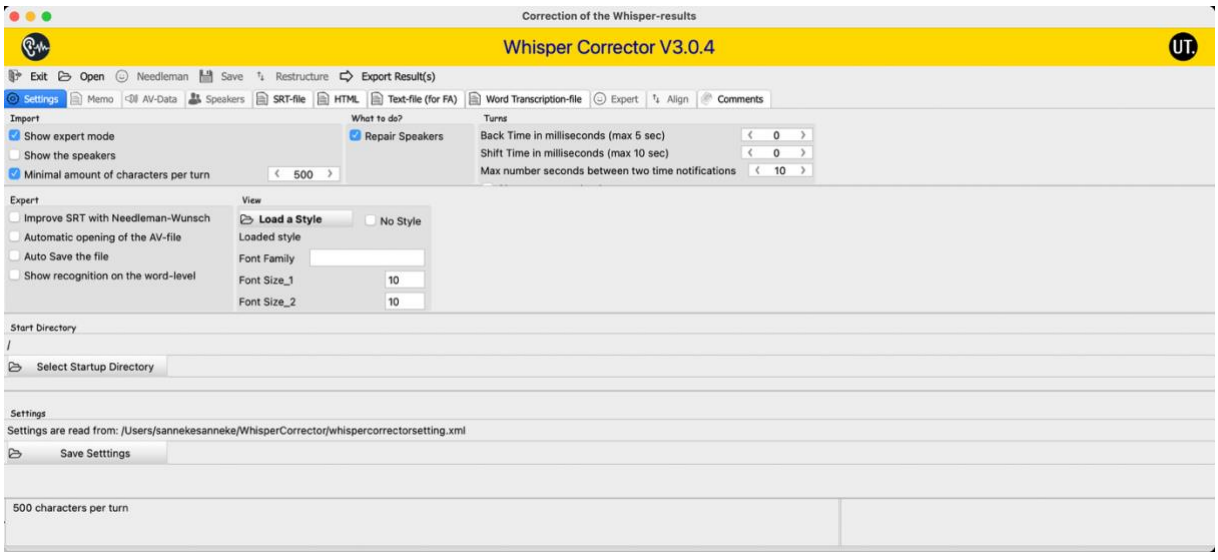


*Fig. 1: WhisperCorrector in Expert mode, de setting waarin je aan het begin het Minimal amount of characters per turn' kunt instellen, de Whisper resultaten in JSON of SRT kunt importeren en na correctie de resultaten kunt exporteren.*

## Open JSON / SRT

In Expert mode, kun je aan het begin de Minimal Amount of Characters per Turns instellen voor het aantal woorden per turn, memoweergave, standaard op 500. Voor een indruk van de hoeveelheid zie Fig. 2.

In Expert mode klik 'open' en importeer eenmalig het *.json-bestand (of: *.srt) uit je folder. Het programma maakt zo een *.xml bestand aan dat is terug te vinden in je folder bij de andere bestanden. Dit is het werkbestand. Je zult dit bewerken in de memoweergave, niet in de Expert mode. Sluit Expert mode.

## Open Audio / Video

Klik 'Open', en importeer het xml bestand. Selecteer het audio- (wav of mp3) of videobestand (mp4). Als je bestandsformaat niet kan gelezen worden, download een WAV-converter en zet het om naar *.wav, ook met video, en werk verder met dit audiobestand. Het totale aantal turns staat links onderaan (Fig.2).

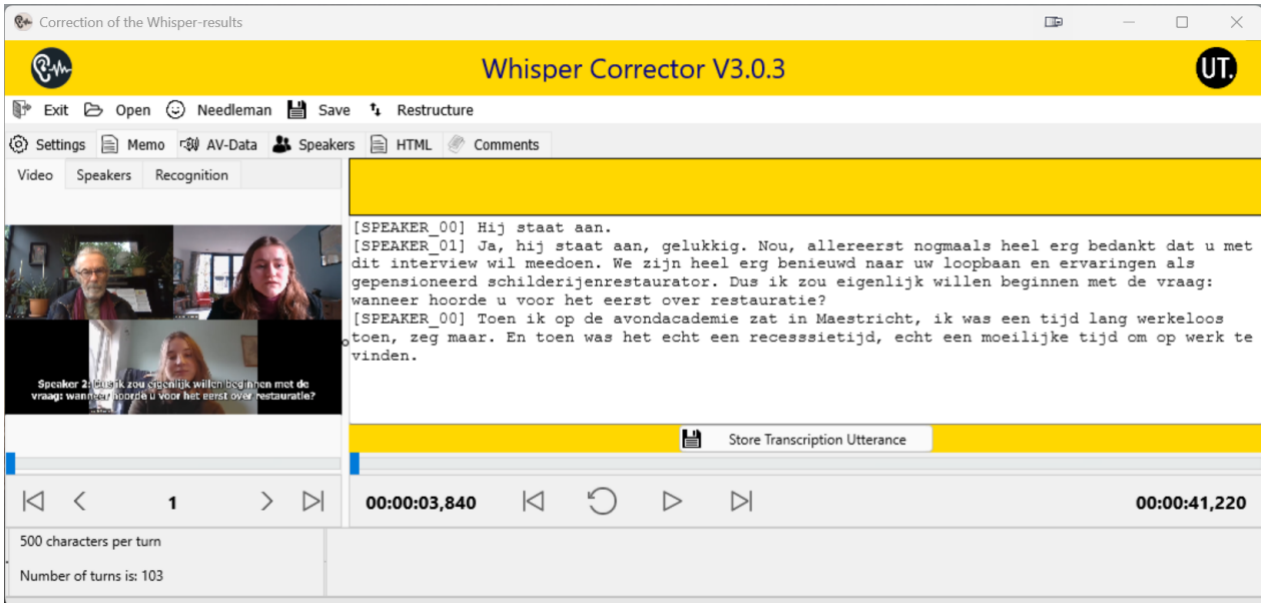


*Fig. 2: Na importeren van een *.json-file en videofile (in die volgorde). Default 500 characters per turn, hier 1 van 103.*

Bij Windows moet je soms de video driver K-Lite Codec Pack Full installeren. Kies op de K-Lite Codec website 'server 1' (of 2 of 3) en installeer de software. Erna leest het programma de video-file goed in.

Na het inlezen klik je op 'Save' en dan 'Save xml'. Het json-bestand heb je nu niet meer nodig. 'Show Expert Mode' is dan default weer uitgezet. Je hebt deze stand nog één keer nodig wanneer je klaar bent, bij het Resultaten exporteren.

### Add Speakers

De laatste stap van de voorbereiding is sprekers toevoegen. Klik op de tab Speakers en vul voor iedere spreker de gegevens in (Fig. 3). Het veld ID wordt automatisch gegenereerd. De velden Gender, Role (Interviewee/Interviewer/Other) en Description zijn optioneel en hebben geen invloed op het correctieproces of de output en kunnen leeggelaten worden. Belangrijk is om het veld Name correct en volledig in te vullen, omdat dat wordt overgenomen in één van de output formats voor het transcript (word.txt). Verbeteren kan achteraf handmatig per turn, waarbij de oude naam niet meer als spreker wordt herkend. Deze komt met vierkante haken tussen de gesproken tekst terecht waar deze handmatig uitgehaald moet worden (vierkante haken mogen verder niet gebruikt in lopende tekst, dat verstoord correcte spreker aanduiding en wisselingen).

Alle sprekers ingevuld? Vergeet niet te klikken op: Replace and Save All Speakers.

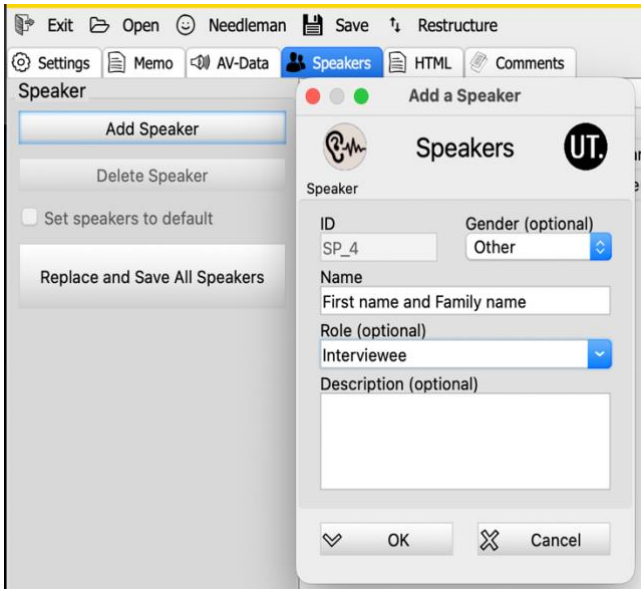


*Fig. 3: Sprekers toevoegen.*

### Save XML

Klik tenslotte helemaal bovenaan op Save, en kies Save XML. Vergeet dit nooit voor je afsluit, anders worden je bewerkingen niet opgeslagen. Tussendoor bewaren kan geen kwaad. Je hebt nu een xml-bestand gemaakt waarin je de automatisch herkende tekst kunt corrigeren.

Nooit meer de *json openen (in Expert mode), tenzij je helemaal opnieuw wil beginnen. Dan wordt namelijk het eerder gemaakte *.xml bestand met je correcties geheel vervangen en begin je weer van voren af aan.

## 2. Corrigeren

Nadat je de srt/json-bestanden hebt ingelezen en in de map met het audio/videobestand hebt opgeslagen als *.xml-bestand, begint de correctiefase van de spraakherkenning. Je opent vanaf nu steeds het xml-bestand – het werkbestand (Fig. 4). Je kunt altijd tussendoor opslaan en WhisperCorrector sluiten en later weer verder gaan. Klik altijd eerst op Save -> Save XML voordat je WhisperCorrector afsluit, om je bewerking te bewaren. De bewerkte xml staat in de map op je computer, en niet in het programma. Je kunt aan meerdere transcripties tegelijk werken met materiaal uit verschillende mappen.

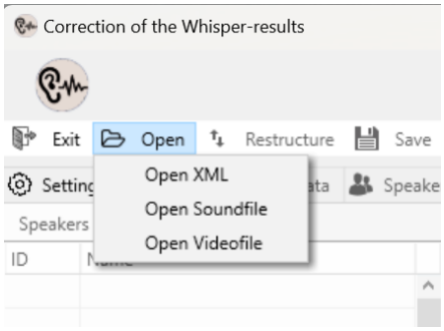


*Fig. 4: Open xml voor het corrigeren*

## Correctie

Klik op Memo voor tekstverwerking van de ASR-resultaten. Rechts staat de herkende tekst waarin je kunt werken. Daaronder audioknoppen en de begin- en eindtijd van het fragment voor het afspelen/pauzeren van het geluidsfragment. De blauwe indicator in de balk er vlak boven kan naar een specifiek moment in het geluidsfragment worden geschoven. Links kun je uit 3 tabs kiezen, Video, Speakers of Recognition. Daaronder knoppen om door het bestand heen te lopen, in grote stappen van -10 'turns' terug of +10 vooruit (buitenste tekens), of in stapjes van -1 of +1 (< of >) (Fig. 5).

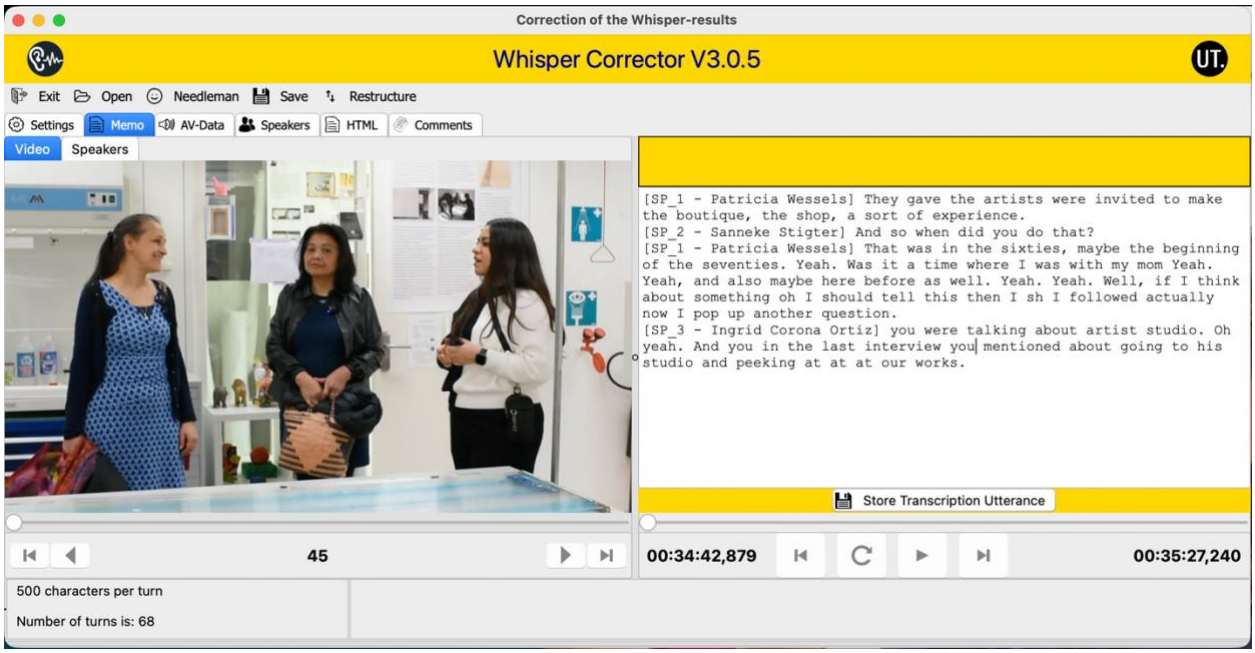


*Fig. 5: Hier kun je horen en lezen wat er gezegd en herkend werd. In het tekstveld kun je corrigeren.*

Je kunt nu de tekst waar nodig verbeteren, interpunctie toevoegen, en eventueel sprekers tussendoor toevoegen (rechtermuisknop), als de herkenner ze niet heeft opgepikt. Harde returns hebben geen betekenis. Voor iedere spreker is een harde return toegevoegd voor de leesbaarheid.

## Turns samenvoegen

Het kan gebeuren dat je wat je hoort wil combineren met de vorige of volgende turn. Dit gebeurt vooral wanneer de beurten vrij klein zijn. Klik daartoe met de rechtermuisknop op het nummer van de beurt in het rechter beneden venster en selecteer Glue turn with prev one / Glue turn with next one (Fig. 6). De 'gelijmde' tekst wordt samengevoegd in één tekstveld. Let op dat je eventuele correcties eerst opslaat, anders gaan die verloren in het lijmproces. Of eerst lijmen, en dan corrigeren.

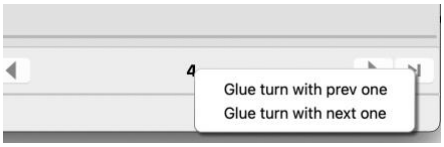


*Fig. 6: Lijmen aan vorige of volgende turn*

## Search & Replace

Als een woord vaker voorkomt en consequent verkeerd is herkand, kan dit in één keer worden vervangen ter correctie. Kan ook voor woord sets. Selecteer het (de) te vervangen woord(en), rechtermuisknop, selecteer Search & Replace en het dialoogvenster opent (Fig. 7). De software kopieert het woord in zowel het zoekveld. Vul het juiste woord in onder 'Replace…' Als je op 'Ok' klikt, wordt het oorspronkelijke woord in het hele document vervangen. Je kunt ook zonder het woord te selecteren zoeken en vervangen door handmatig invoeren.

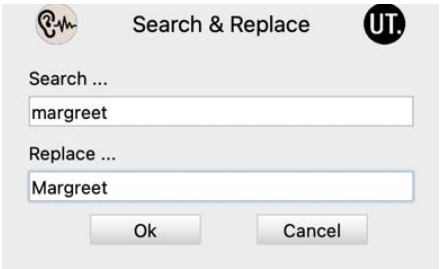


*Fig. 7: Dialoogvenster voor Search & Replace. Ook handig om naderhand typo's corrigeren.*

# 3. Resultaten exporteren

Wanneer je klaar bent met corrigeren van de herkenningsresultaten, kun je de verbeterde tekst exporteren en opslaan voor ondertiteling (SRT), Word-transcript en als “schone” tekstfile. Dat gaat als volgt:

1. Open gecorrigeerde xml, kies Settings
2. Selecteer ‘Show Expert mode’;
3. Klik op Export Results in format naar keuze, of All files in één keer (Fig. 8).

De resultaten belanden in de map met alle andere bestanden met dezelfde bestandsnaam.

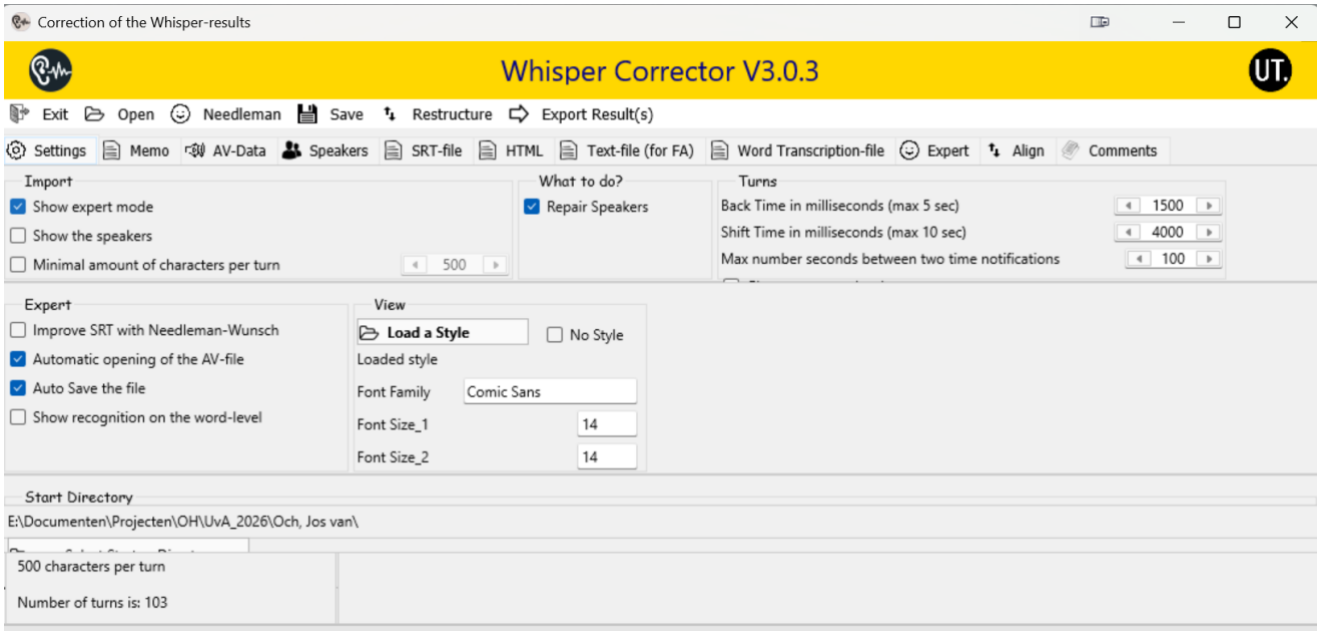


*Fig. 8. Exporteren van de gecorrigeerde resultaten*

### Uitleg

- *.srt, een ondertitelingsformaat, overschrijft het oorspronkelijke *.srt bestand met de gecorrigeerde versie. Deze kan automatisch meespelen wanneer deze in dezelfde map als de video staat (totdat je exporteert, kan het zijn dat je ongecorrigeerde tekst in beeld ziet meelopen). Kan makkelijk worden opgezet in andere ondertitelingstypen, zoals *.vtt;
- *.html, een doorzoekbare audio-tekst interface, functioneert wanneer in zelfde map met AV-bestand;
- *.txt, een rekenkundig goed leesbaar tekstbestand, overschrijft het oorspronkelijke txt-bestand. Voorzien van tijdscodes;
- *.word.txt, met vooringestelde opmaak. Deze kun je openen, de inhoud kopiëren en plakken in Word voor leesbaar transcript met sprekers voor de tekst, door linker kantlijn 3 cm naar rechts schuiven.
- Bewaar alle bestanden bij elkaar in dezelfde map, zodat ze in programma’s tegelijkertijd in te lezen zijn.

### Tips

1. Typos worden in de *.txt zichtbaar omdat ze rood onderstreept zijn. Je kunt dit ter controle gebruiken en in de WhisperCorrector de laatste correcties aanbrengen met ‘Search & Replace’.
2. Archiveer ook deze gecorrigeerde *.txt- en *.xml (werk) bestanden. Deze bestandsformaten blijven naar verwachting goed leesbaar en zijn bruikbaar voor verschillende taalanalyse tools (interoperabel).
3. Maak van je leesbare transcript met metadata ervoor een pdf. In eventueel een tweede versie kun je het transcript annoteren, bijvoorbeeld naar aanleiding van input van de geïnterviewde en duidelijk maakt in noten, eventueel binnen vierkante haken in de tekst.

# Good practice

Enkele suggesties om ervoor te zorgen dat de transcripties zo goed en uniform mogelijk worden uitgevoerd.

### Zinnen

Onvolledige zinnen eindigen met drie puntjes aan het eind.

Bijvoorbeeld:

> "En toen dacht ik..."

Met die drie puntjes geef je aan dat de zin eindigt, maar dat er eigenlijk een deel van de tekst ontbreekt.

Zorg ervoor dat de zinnen niet te lang zijn, ook al is één lange zin taalkundig correct. Voor de leesbaarheid, verdient het de voorkeur om een lange zin in kortere zinnen op te splitsen.

Bijvoorbeeld:

> "We zijn er nog elke dag mee bezig en dat wil ik de mensen ook meegeven dat de meeste mensen die hier nu naar gaan kijken het ook met mij eens zullen zijn dat het nooit af is."

Leest beter als volgt:

> "We zijn er nog elke dag mee bezig. En dat is ook wat ik de mensen wil meegeven, dat de meeste mensen die er nu naar gaan kijken, het ook met mij eens zullen zijn dat het nooit af is."

### Herhaling

Whisper laat gebruik van 'ehms' meestal automatisch weg tenzij ze heel nadrukkelijk worden uitgesproken. Vermijd ook onnodige herhaling. Veelvuldige herhaling kun je bijvoorbeeld beperken tot 3 maal, wat prettiger leest en waaruit de repetitie evengoed blijkt.

Bijvoorbeeld:

In plaats van "Ja, ja, ja, ja, ja, ja", noteer je: "Ja, ja, ja."

### Interpunctie

Met goed en zorgvuldig gebruik van interpunctie is gesproken tekst goed leesbaar te maken, dus besteedt hier voldoende aandacht aan.

Regels voor interpunctie verschillen per taal. Het Nederlands kent andere regels dan het Engels. Voor Engels, zie: The Punctuation Guide.

- Gebruik geen losse interpunctietekens, zoals [spatie] – [spatie]. Een zwevend leesteken zal als woord worden opgevat bij het uitlijnen, waardoor tekst en audio niet meer synchroniseren.
- Dus: een komma zit altijd 'vast' aan het woord, zoals hier, evenals 'aanhalingstekens' en punten. Inclusief afloop puntjes, overigens…
- Na aflooppuntjes geen volgend leesteken gebruiken, zoals een komma of aanhalingsteken. De zin is met 3 puntjes beëindigd.
- Gebruik géén voorlooppuntjes.
- Gebruik geen vierkante haken in de tekst […], die zijn voor de werking van het programma. Je vindt de sprekers terug in de spreker weergave in het uiteindelijke transcript (word.txt), terwijl de namen juist weggelaten worden in de ondertiteling (srt).