

## Standoff-EXMARaLDA-Transcription Annotation Tool (Sextant)

---

Sextant ist eine *proof-of-concept*-Implementation für das System der EXMARaLDA-Standoff-Annotationen. Es dient zur Validierung und einfachen Visualisierung von EXMARaLDA-Standoff (.esa)-Dateien.

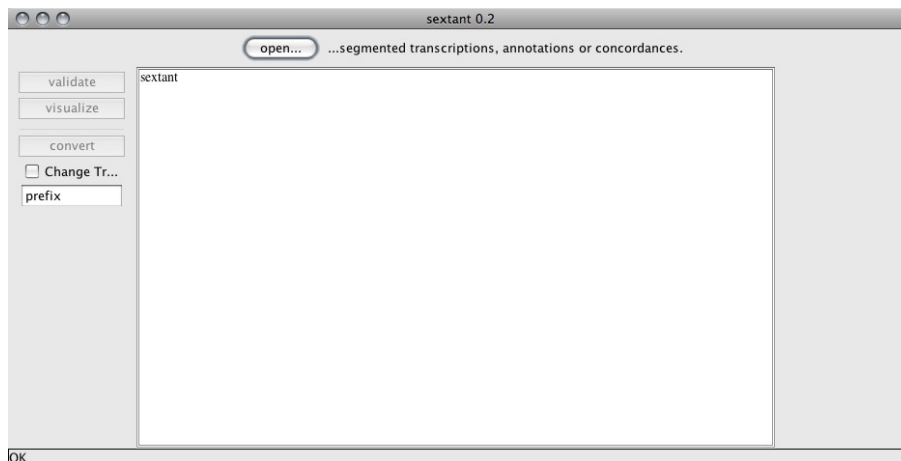
Dieses Dokument dient als kurze Einführung in das Programm – es ist kein ausführliches Manual.

### Starten des Programms

Um das Programm zu starten benötigen Sie eine funktionierende Java Laufzeitumgebung der Version 6 oder höher. Wie sie diese bekommen können erfahren Sie z.B. unter <http://www.exmaralda.org/faq.html>.

Das Programm kann als ausführbares Java-Paket unter <http://www.exmaralda.org/sextant> heruntergeladen werden. Wenn Sie Java richtig installiert haben, startet ein Doppelklick auf das Programmsymbol die Anwendung.

Sie sollten ungefähr den folgenden Bildschirm sehen:



### Das Programm benutzen

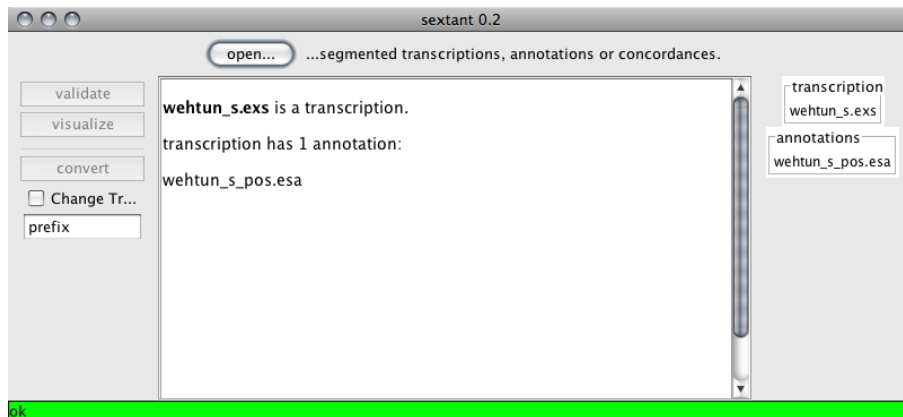
Mit dem Programm können Sie drei verschiedene Dateitypen öffnen:

- Segmentierte EXMARaLDA-Transkriptionen (.exs oder .xml-Dateien)
- EXMARaLDA-Standoff-Annotationen (.esa-Dateien)
- EXMARaLDA Konkordanzen (.exc oder .xml-Dateien)

Um eine Datei zu öffnen, klicken Sie auf die „open...“-Schaltfläche am oberen Rand des Bildschirms.

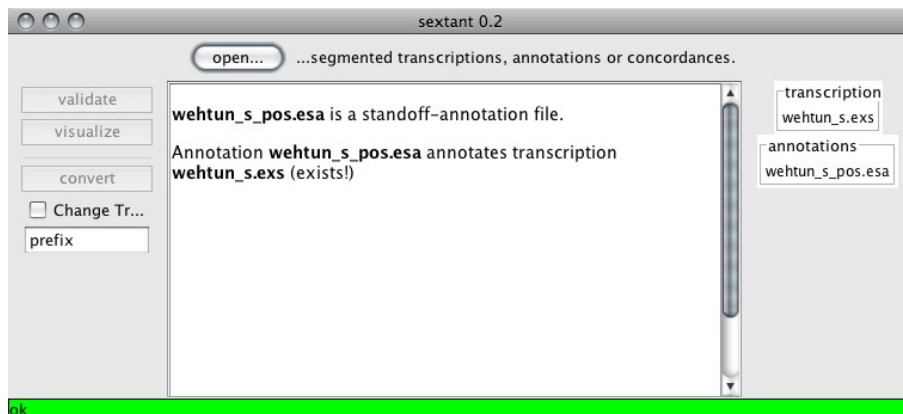
Je nach geöffnetem Dateityp bietet das Programm verschiedene Möglichkeiten:

Wenn Sie eine **EXMARaLDA-Transkription** öffnen, so sollte der Bildschirm etwa so aussehen:



Es wird angezeigt, dass der Transkription eine Annotationsdatei zugeordnet ist („wehtun\_s\_pos.esa“). Auf der rechten Seite sehen Sie in der Tabelle „transcriptions“ die geladene Transkriptionsdatei und in der Tabelle „annotations“ die zugeordnete Annotationsdatei.

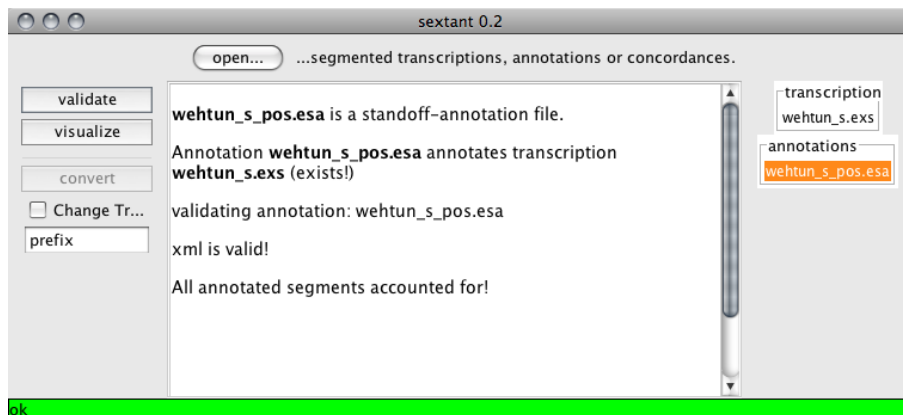
Sie erhielten etwa den gleichen Bildschirm, wenn Sie die entsprechende **Standoff-Annotationsdatei** geöffnet hätten:



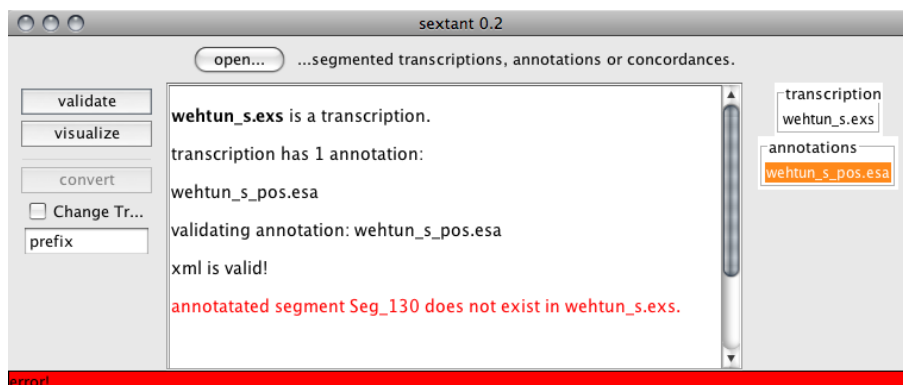
Sie können nun eine der beiden Dateien am rechten Rand anklicken, um die auf der linken Seite verfügbaren Schaltflächen für diesen Dateityp zu aktivieren.

## Validieren

Wenn Sie die Annotationsdatei anklicken, so werden die Schaltflächen „validate“ und „visualize“ aktiviert. Klicken Sie die Schaltfläche „validate“ um zu validieren, ob die Annotationsdatei valide ist und alle annotierten Segmente auch tatsächlich in der Transkriptionsdatei existieren. In diesem Fall sollten Sie folgende Rückmeldung erhalten:



Es können aber auch Fehler bei der Validierung auftreten. Das könnte z.B. so aussehen:



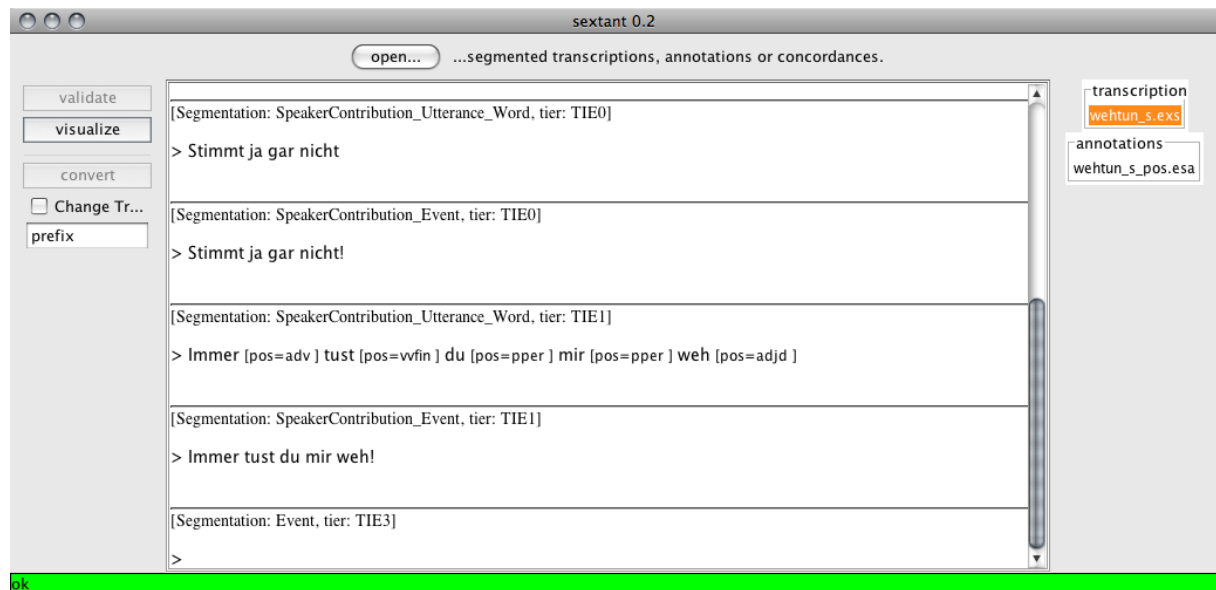
In diesem Fall existiert ein Segment, für das eine Annotation vorliegt, nicht in der Transkription.

Beachten Sie, dass Sextant keine Standoff-Annotationen validieren kann, die mit XPointern auf Teilsegmente in der Transkription zeigen. Das Programm gibt in einem solchen Fall eine entsprechende Meldung aus.

## Visualisieren

Sowohl für Annotationen als auch für Transkriptionen (sofern diese über Annotationen verfügen) wird die Schaltfläche „visualize“ aktiviert. Wenn Sie diese Schaltfläche

drücken, wird die Transkription mit der (ersten) zugeordneten Annotationsdatei im Programmfenster visualisiert. Das kann in etwa so aussehen:

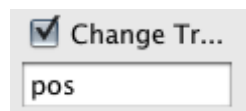


Sextant geht in diesem Fall alle in der Transkription enthaltenen Segmentierungen der Reihe nach durch und gibt die Einzelsegmente zusammen mit ihrer zugeordneten Annotation aus. Im Beispiel sind nur Segmente auf der Ebene „SpeakerContribution\_Utterance\_Word“ annotiert, deshalb finden sich auch nur auf dieser Ebene Annotationen (in eckigen Klammern).

### Konvertieren:

Wenn Sie eine Konkordanz aus dem EXMARaLDA-Suchwerkzeug „EXAKT“ geöffnet haben, so wird die Schaltfläche „convert“ aktiviert. Mit einem Druck auf diese Schaltfläche werden Annotationen, die einer Ergebnisliste des Suchwerkzeugs vorgenommen wurden, in Standoff-Annotationen der durchsuchten EXMARaLDA-Transkriptionen konvertiert.

Bevor Sie die Schaltfläche drücken können Sie noch zwei Optionen einstellen:



Wenn Sie die Checkbox „Change Tr...(anscription)“ anklicken, so werden die neu angelegten Standoff-Annotationsdateien in den Metadaten der Transkriptionen vermerkt, auf die sich die Annotationen beziehen. Im Eingabefeld darunter können Sie ein Suffix eintragen, mit dem die Dateinamen der Annotationsdateien nach dem Muster

[Dateiname der Transkriptionsdatei]\_[Suffix].esa

gebildet werden. Die Annotationsdateien werden immer in das selbe Verzeichnis geschrieben wie die annotierten Transkriptionsdateien.

Wenn Sie eine Konkordanz geöffnet, die entsprechenden Einstellungen vorgenommen und die „convert“-Schaltfläche gedrückt haben, wird der Konvertierungsprozess gestartet:

```
cosiconcordance.xml is a concordance.
looking for annotated transcription files...
/Volumes/NW1000_SFB538B/Daten/TP-Z2/DATEN/K6/1.0/Berlin/Berlin_s.xml has annotations
/Volumes/NW1000_SFB538B/Daten/TP-Z2/DATEN/K6/1.0/Heidelberg/Heidelberg_10_s.xml has annotations
/Volumes/NW1000_SFB538B/Daten/TP-Z2/DATEN/K6/1.0/Hamburg/Hamburg_s.xml has annotations

converting annotations and changing transcription headers

annotation types found:
POS (ClosedCategoryList)
JaNein (Binary)

/Volumes/NW1000_SFB538B/Daten/TP-Z2/DATEN/K6/1.0/Berlin/Berlin_s_pos.esa written.
changed transcription head in Berlin_s.xml
/Volumes/NW1000_SFB538B/Daten/TP-Z2/DATEN/K6/1.0/Heidelberg/Heidelberg_10_s_pos.esa written.
changed transcription head in Heidelberg_10_s.xml
/Volumes/NW1000_SFB538B/Daten/TP-Z2/DATEN/K6/1.0/Hamburg/Hamburg_s_pos.esa written.
changed transcription head in Hamburg_s.xml
...done!
```

In der Ausgabe werden zunächst die Transkriptionsdateien aufgelistet, für die Annotationen vorliegen. Dann werden die Annotationstypen aufgelistet, die in der Konkordanz gefunden wurden. Daraufhin wird die Konvertierung gestartet und die geschriebenen bzw. modifizierten Dateien angegeben.

Beachten Sie, dass Sextant die geschriebenen Annotationsdateien nicht validieren und visualisieren kann, da sie auf XPointer-Verweisen basieren.