

# CorA (Corpus Annotator) Benutzerhandbuch

Version 1.0

Marcel Bollmann  
bollmann@linguistics.rub.de  
Sprachwissenschaftliches Institut  
Ruhr-Universität Bochum

10. Juli 2013

## Inhaltsverzeichnis

<b>1</b>	<b>Allgemeines</b>	<b>3</b>
1.1	Benutzername und Passwort . . . . .	3
<b>2</b>	<b>Dateiverwaltung</b>	<b>4</b>
2.1	Öffnen einer Datei . . . . .	4
2.2	Importieren von Transkriptionen . . . . .	5
2.3	Exportieren . . . . .	6
2.4	Löschen . . . . .	6
<b>3</b>	<b>Das Editor-Fenster</b>	<b>7</b>
3.1	Navigationsleiste . . . . .	7
3.2	Spalten im Editor . . . . .	8
3.2.1	Fortschrittsbalken (P) . . . . .	8
3.2.2	Nummerierung . . . . .	8
3.2.3	Markierung von Problemfällen (E) . . . . .	9
3.2.4	Token . . . . .	9
3.2.5	POS- und Morphologie-Tags . . . . .	9
3.2.6	Lemmatisierung . . . . .	10
3.2.7	Kommentarfeld . . . . .	11
3.2.8	Kontextmenü . . . . .	11
3.3	Horizontale Textansicht . . . . .	11
3.4	Anpassen des Editors . . . . .	11
<b>4</b>	<b>Bearbeiten der Transkription</b>	<b>13</b>
4.1	Token bearbeiten . . . . .	13
4.2	Token hinzufügen . . . . .	14

4.3	Token löschen . . . . .	15
4.4	Praktische Beispiele . . . . .	15
<b>5</b>	<b>Tastaturbefehle</b>	<b>17</b>
<b>6</b>	<b>Umgang mit Fehlermeldungen</b>	<b>18</b>

# 1 Allgemeines

CorA ist ein web-basiertes Programm zur Annotation von Korpora. Die aktuelle Adresse (URL) von CorA lautet:

`http://smokehead.linguistics.rub.de/cora/`

Wir empfehlen die Verwendung einer aktuellen Version von Google Chrome<sup>1</sup> (bzw. Chromium) oder Mozilla Firefox<sup>2</sup>. JavaScript darf nicht deaktiviert sein. CorA funktioniert möglicherweise auch mit anderen Browsern, jedoch wird das Tool von uns nur mit den genannten Browsern getestet.

## Wichtig!

Die Benutzung von CorA erfordert eine ständige Internetverbindung. Es ist nicht möglich, das Tool offline zu benutzen. Wenn Sie während der Bearbeitung eines Dokuments länger als 30 Minuten offline sind, kann ein Datenverlust nicht ausgeschlossen werden.

Bitte öffnen Sie CorA **nicht** gleichzeitig in mehreren Browserfenstern auf demselben Rechner, und loggen Sie sich **nicht** gleichzeitig mit demselben Benutzerkonto auf mehreren Rechnern ein. Dies kann zu unvorhergesehenem Verhalten führen, da CorA hierfür nicht ausgelegt ist.

## 1.1 Benutzername und Passwort

Bevor Sie CorA benutzen können, müssen Sie sich mit Benutzernamen und Passwort anmelden. Falls ein neues Benutzerkonto angelegt werden soll, oder Sie Ihr Passwort vergessen haben, wenden Sie sich bitte direkt an Marcel Bollmann ([bollmann@linguistics.rub.de](mailto:bollmann@linguistics.rub.de)).

Ihr **Passwort ändern** können Sie hingegen direkt in CorA: hierzu loggen Sie sich zunächst ein, wechseln zum Reiter „Einstellungen“ und klicken dort auf die Schaltfläche „Passwort ändern...“. Wenn Sie sich zum ersten Mal einloggen, sollten Sie als erstes unbedingt Ihr Passwort ändern!

---

<sup>1</sup><http://www.google.com/chrome/>

<sup>2</sup><http://www.mozilla.com/>

## 2 Dateiverwaltung

Nach der Anmeldung in CorA wird zunächst der Reiter „Datei“ angezeigt. Abbildung 1 zeigt einen Beispiel-Screenshot, der im Folgenden näher erläutert wird.

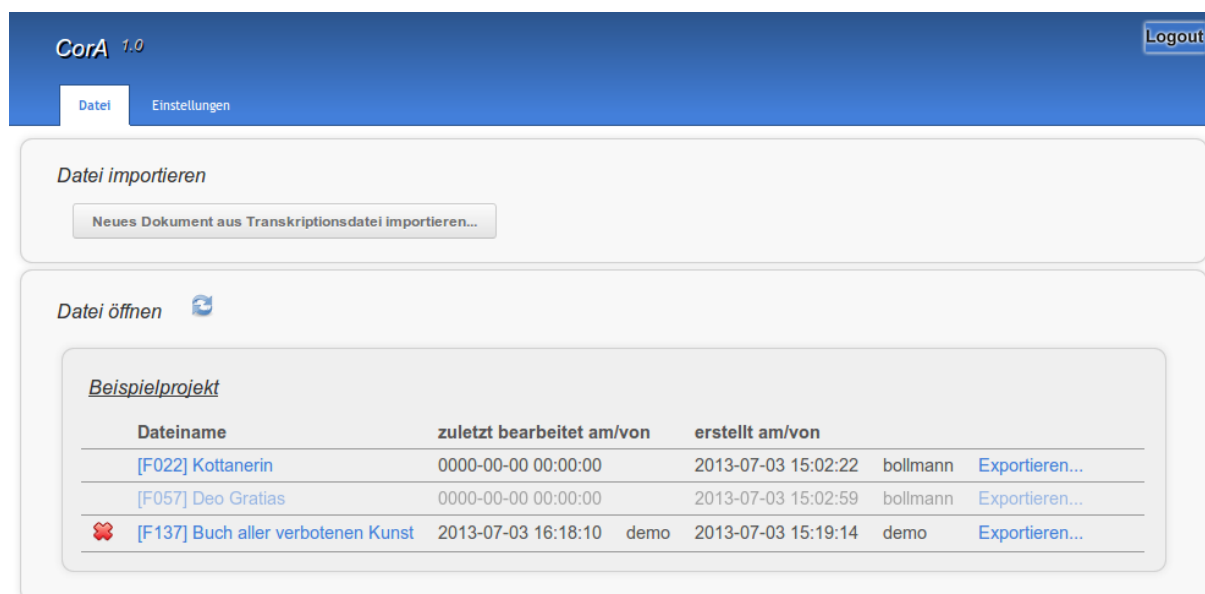
### 2.1 Öffnen einer Datei

Klicken Sie einfach auf einen Dateinamen, um die entsprechende Datei zum Annotieren zu öffnen. Nun öffnet sich automatisch das Editor-Fenster (s. Abschnitt 3).

Dateinamen, die in der Liste **leicht ausgegraut** erscheinen (wie z.B. „Deo Gratias“ in Abbildung 1), sind bereits von einem anderen Nutzer geöffnet. Eine Datei kann immer **nur von einem Nutzer gleichzeitig** geöffnet sein! Klicken Sie den Dateinamen an, um zu erfahren, wer die Datei aktuell geöffnet hat. Die Datei kann erst wieder von jemand anderem geöffnet werden, wenn dieser Nutzer entweder

- die Datei über den Knopf „Datei schließen“ schließt,
- sich aus CorA ausloggt (über den Knopf „Logout“ oben rechts), oder
- länger als 30 Minuten nicht mehr online gewesen ist.

Durch Klicken auf das Symbol  wird die Dateiliste aktualisiert.




**CorA 1.0** Logout

Datei Einstellungen

**Datei importieren**

Neues Dokument aus Transkriptionsdatei importieren...

**Datei öffnen** 

Beispielprojekt


Dateiname	zuletzt bearbeitet am/von		erstellt am/von		
[F022] Kottanerin	0000-00-00 00:00:00		2013-07-03 15:02:22	bollmann	Exportieren...
[F057] Deo Gratias	0000-00-00 00:00:00		2013-07-03 15:02:59	bollmann	Exportieren...
 [F137] Buch aller verbotenen Kunst	2013-07-03 16:18:10	demo	2013-07-03 15:19:14	demo	Exportieren...

Abbildung 1: Dateien verwalten

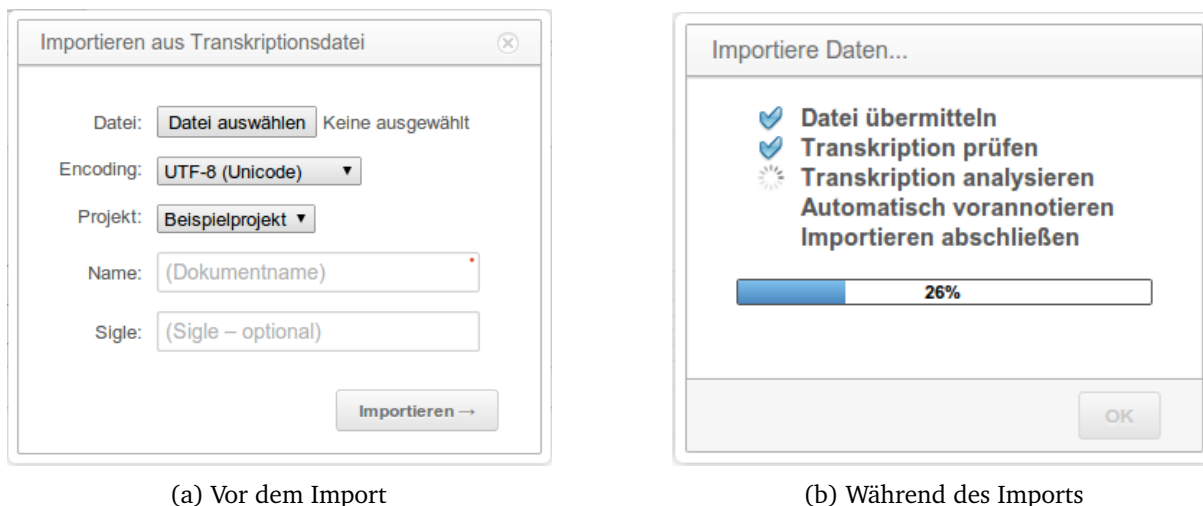


Abbildung 2: Importieren einer Datei

## 2.2 Importieren von Transkriptionen

Um eine neue Datei hinzuzufügen, klicken Sie auf den Knopf „Neues Dokument aus Transkriptionsdatei importieren...“ (s. Abbildung 1).

Es öffnet sich ein Fenster ähnlich wie in Abbildung 2a. Wählen Sie hier die **Datei** aus, die Sie hochladen möchten, und geben Sie an in welchem **Encoding** die Datei gespeichert wurde. Stellen Sie ggf. sicher, dass das richtige **Projekt** ausgewählt ist (z.B. „Referenzkorpus Frühneuhochdeutsch“). Sie müssen außerdem **Name** und **Sigle** des Dokuments angeben, unter denen es in CorA geführt werden soll.

Der Import startet nach Klick auf „Importieren“. Es öffnet sich ein neues Fenster, welches den Fortschritt des Importvorgangs anzeigt (Abbildung 2b). Je nach Umfang der Datei kann der Import **ca. 10 bis 15 Minuten** in Anspruch nehmen. Bitte schließen Sie das Browserfenster in dieser Zeit nicht!<sup>3</sup> Arbeiten Sie auch nicht in einem anderen Fenster mit CorA weiter!

### Fehlermeldungen


Während des Imports können u.U. auch Fehlermeldungen auftreten, z.B. wenn das Check-Skript noch Fehler in der Transkription bemängelt oder wenn das Encoding falsch angegeben wurde. Falls jedoch andere Fehlermeldungen mit für Sie unverständlichem Inhalt auftreten sollten, melden Sie uns diese und beachten bitte unbedingt auch die Hinweise in Abschnitt 6, die wir extra für diesen Fall zusammengestellt haben!

<sup>3</sup>Wenn Sie das Browserfenster während eines laufenden Imports schließen, läuft der Import trotzdem weiter. Sie können allerdings möglicherweise auftretende Fehlermeldungen nicht mehr sehen, und eventuell auch nicht mehr auf CorA zugreifen, bevor der Import beendet ist.

## 2.3 Exportieren

Jede Datei in der Dateiliste besitzt am Ende der Zeile einen Link „Exportieren. . .“ (s. Abbildung 1). Diese Funktionalität ist in der aktuellen CorA-Version jedoch noch nicht implementiert und wird zu einem späteren Zeitpunkt nachgeliefert.

## 2.4 Löschen

Dateien können durch Klick auf das rote Kreuz  vor dem jeweiligen Dateinamen (s. Abbildung 1) wieder gelöscht werden. Das Löschen von Dateien ist endgültig und **kann nicht rückgängig gemacht werden!** Es erscheint daher eine entsprechende Warnung, die bestätigt werden muss, bevor die Datei tatsächlich gelöscht wird.

Nur der Benutzer, der die Datei importiert hat, kann sie auch wieder löschen. Allen anderen Nutzern wird das entsprechende Symbol gar nicht erst angezeigt.

### 3 Das Editor-Fenster

Der Editor ist der zentrale Bestandteil von CorA. Er öffnet sich automatisch, wenn eine Datei geöffnet wird, bzw. durch Klick auf den Reiter „Editor“ am oberen Bildschirmrand. Abbildung 3 zeigt einen Beispiel-Screenshot.

#### 3.1 Navigationsleiste

Am oberen Rand des Editors befindet sich eine Leiste, mit der innerhalb des Dokuments navigiert werden kann:

Seite: |<< < 1 ▾ > >>| [Springe zu Zeile...](#)

- „|<<“ bzw. „|>>“ springt zur ersten bzw. letzten Seite des Dokuments
- „<“ bzw. „>“ blättert eine Seite zurück bzw. vor
- Die Dropdown-Box in der Mitte zeigt die momentane Seitenzahl an; Auswählen einer anderen Seitenzahl springt zur entsprechenden Seite.
- Klicken auf „Springe zu Zeile...“ öffnet ein Fenster, wo eine Zeilennummer eingegeben werden kann; der Editor springt dann auf die Seite, die die entsprechende Zeilennummer enthält.

Zeilen- und Seitenzahlen beziehen sich dabei immer auf die Darstellung im Editor, **nicht** auf Zeilen bzw. Seiten in der Transkription! Funktionen zum Navigieren anhand der Nummerierung in der Original-Transkription sind momentan noch nicht implementiert.

The screenshot shows the CorA 1.0 Editor interface. At the top, there's a blue header bar with the title 'CorA 1.0 Buch aller verbotenen Kunst' and a 'Logout' button. Below the header, there are tabs for 'Datei', 'Editor', and 'Einstellungen'. On the right, there are buttons for 'Datei speichern' and 'Datei schließen'. The main area displays a table with the following columns: P, #, Zeile, E, Token (Trans), Token (UTF), POS-Tag, Morphologie-Tag, Lemma, and Kommentar. The table contains 14 rows of data, with some cells highlighted in red. The navigation bar at the top of the editor shows 'Seite: |<< < 1 ▾ > >>| Springe zu Zeile...'.

P	#	Zeile	E	Token (Trans)	Token (UTF)	POS-Tag	Morphologie-Tag	Lemma	Kommentar
1	01r,01	<input type="checkbox"/>	Die	Die	DARTB	Fem.Nom.Sg	<input type="checkbox"/>	d	
2	01r,01	<input type="checkbox"/>	vor#red	vorred	NA	Fem.Nom.Sg	<input type="checkbox"/>	vorrede [GV14469]	
3	01r,01	<input type="checkbox"/>	in	in	APPR	–	<input type="checkbox"/>		
4	01r,01	<input type="checkbox"/>	dem	dem	?	–	<input type="checkbox"/>		
5	01r,01	<input type="checkbox"/>	puloch	püch	?	–	<input type="checkbox"/>		
6	01r,01	<input type="checkbox"/>	aller	aller	?	–	<input type="checkbox"/>		
7	01r,02	<input type="checkbox"/>	verpoten	verpoten	VVPPA	*.*.*	<input type="checkbox"/>		
8	01r,02	<input type="checkbox"/>	kun\$	kunft	NA	Fem.Dat.Sg	<input type="checkbox"/>		
9	01r,02	<input type="checkbox"/>	(.)	(.)	\$.	–	<input type="checkbox"/>		
10	01r,02	<input type="checkbox"/>	vngelaubens	vngelaubens	?	–	<input type="checkbox"/>		
11	01r,02	<input type="checkbox"/>	vnd	vnd	?	–	<input type="checkbox"/>		
12	01r,03	<input type="checkbox"/>	der	der	?	–	<input type="checkbox"/>		
13	01r,03	<input type="checkbox"/>	zaubrey	zaubrey	NA	*.Nom.Sg	<input type="checkbox"/>		
14	01r,03	<input type="checkbox"/>	(.)	(.)	\$.	–	<input type="checkbox"/>		

Abbildung 3: Das Editor-Fenster

## 3.2 Spalten im Editor

Wortformen im CorA-Editor werden immer zeilenweise angezeigt, wobei jede Zeile ein Token nach moderner Tokenisierung enthält. Im Folgenden werden Bedeutung und Funktionsweise der einzelnen Spalten erläutert.

### 3.2.1 Fortschrittsbalken (P)

Der Fortschrittsbalken soll anzeigen, bis zu welchem Punkt eine Datei bereits bearbeitet worden ist. Eine **grüne** Markierung zeigt den bereits bearbeiteten Bereich an, während der Balken ansonsten **grau** hinterlegt ist.



Diese Markierung erfüllt derzeit folgende Funktionen:

- Wenn eine Datei geöffnet wird, springt der Editor automatisch an die Stelle, wo der Fortschrittsbalken endet.
- Die automatische Lernfunktion lernt nur aus den Daten, die bereits grün markiert sind. (noch nicht implementiert)

Wird eine Annotation an einem Token geändert, verlängert sich der Fortschrittsbalken ggf. automatisch bis zu diesem Punkt. Bedenken Sie dies insbesondere, falls Sie einmal an eine spätere Stelle im Text springen, um dort eine Annotation zu ändern! Sie müssen sich in diesem Fall merken, an welchem Punkt Sie zuvor gewesen sind, da die Fortschrittsanzeige sich automatisch verlängert.

Durch Klick auf den Fortschrittsbalken kann dieser allerdings bei Bedarf auch manuell eingestellt werden: Klick auf einen grünen Bereich verkürzt den Balken bis zu diesem Punkt, Klick auf einen grauen Bereich verlängert ihn entsprechend.

### 3.2.2 Nummerierung

Zwei verschiedene Zeilennummern werden angezeigt:

1. Die interne Nummerierung (Spalte „#“) ist eine fortlaufende Zählung aller Token; über den Link „Springe zu Zeile. . .“ in der Navigationsleiste kann z.B. direkt zu einer bestimmten Zeilennummer gesprungen werden.
2. Die Spalte „Zeile“ gibt die Zeilennummer in der Original-Transkription an. (In Einzelfällen, wenn ein Wort sich über mehrere Zeilen erstreckt, kann diese Angabe ungenau sein. Dies ist eine bekannte Einschränkung.)

#	Zeile
176	01v,07
177	01v,07
178	01v,08



### 3.2.3 Markierung von Problemfällen (E)

Die Spalte „E“ (ursprünglich für „Error“) enthält eine Checkbox, in der durch Anklicken eine rote Markierung gesetzt bzw. entfernt werden kann. Diese Markierung ist lediglich eine Hilfe für den Bearbeiter und hat intern keine weitere Funktion. Sie kann z.B. benutzt werden, um Fehler oder Problemfälle zu markieren, die später noch einmal angesehen werden müssen.

E
<input type="checkbox"/>
<input checked="" type="checkbox"/>
<input type="checkbox"/>

Es wird in Zukunft eine Suchfunktion geben, mit der gezielt nach derart markierten Token gesucht werden kann.

### 3.2.4 Token

Es gibt zwei verschiedene Darstellungen der Wortformen, die annotiert werden sollen: **Token (Trans)** zeigt die Original-Transkription, während **Token (UTF)** eine Unicode-Darstellung der Transkriptionszeichen enthält. Es ist möglich, den Editor so anzupassen, dass nur eine der beiden Formen dargestellt wird – siehe dazu Abschnitt 3.4.

### 3.2.5 POS- und Morphologie-Tags

Die Wortarten-Annotation in CorA erfolgt in zwei Schritten: zunächst sollte über die entsprechende Dropdown-Box ein **POS-Tag** ausgewählt werden; anschließend kann in der Spalte **Morphologie-Tag** eine Kombination von morphologischen Angaben ausgewählt werden.

POS-Tag	Morphologie-Tag
NA	Neut.Nom.Sg
NA	–
Alle Tags	–
S(	Pos.Fem.Dat.Sg
S,	Fem.Dat.Sg
S.	Mask.Gen.Sg
\$QL	Pos.Mask.Gen.Sg
\$QR	
\$	
ADJA	

Die Auswahlmöglichkeiten bei den Morphologie-Tags sind dabei **abhängig vom gewählten POS-Tag**. Falls der POS-Tag geändert wird, und der aktuell selektierte Morphologie-Tag keine zulässige Auswahl mehr für den neuen POS-Tag ist, so ändert CorA den Morphologie-Tag **automatisch** auf einen zulässigen Wert. Dieser wird natürlich in der Regel noch falsch sein! Es ist daher empfehlenswert, die beiden Spalten immer zusammen zu bearbeiten, bzw. nach Änderung des POS-Tags auch immer die morphologische Angabe zu überprüfen.

Manchmal können die Dropdown-Boxen **rot umrandet** sein. Dies bedeutet, dass der entsprechende Tag auf jeden Fall fehlerhaft ist und geändert werden sollte (z.B., wenn der Tagger nur „?“ als Tag vorgeschlagen hat).

### 3.2.6 Lemmatisierung


Die Lemma-Spalte hat drei Bestandteile: eine **Checkbox**, um Lemma-Einträge als „bestätigt“ zu markieren; ein **Eingabefeld** für das Lemma; und einen **Verweis** auf die Online-Ausgabe des Grimmschen Wörterbuchs.

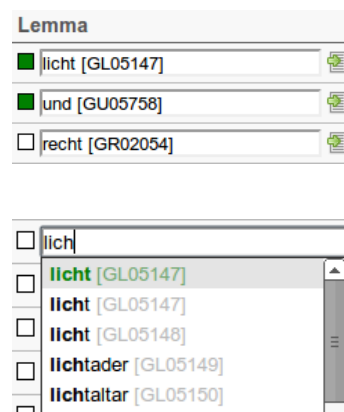
Das Eingabefeld ist mit einer **Dropdown-Liste mit Lemma-Vorschlägen** verknüpft. Diese Liste öffnet sich automatisch, sobald eine Eingabe in dem Feld gemacht wird (s. Abbildung rechts). Alternativ lässt sie sich auch durch Doppelklick in das Eingabefeld oder mit „Pfeiltaste nach unten“ öffnen. Die Liste kann verschiedene Arten von Einträgen enthalten:

1. **Grau hinterlegte** Einträge sind automatisch generierte Vorschläge.
2. Zusätzlich **grün hervorgehobene** Einträge sind Lemmata, die an anderer Stelle für eine gleichlautende Wortform bereits eingetragen und als „bestätigt“ markiert wurden. Diese Vorschläge werden aus allen Texten im selben Projekt generiert. (Es ist allerdings erforderlich, dass die Lemma-Einträge bereits gespeichert wurden, damit sie für andere gleichlautende Wortformen als Vorschlag angezeigt werden können.)
3. Alle anderen Einträge sind durch Autovervollständigung gefundene Treffer im Grimmschen Wörterbuch. Dabei gilt:
  - a) es wird immer nur nach passendem Wortanfang gesucht;
  - b) Diakritika in Grimm werden ignoriert, d.h. Eingabe von *geta* findet sowohl *getacht* wie auch *getäckel* oder *getâkel*; und
  - c) wenn es sehr viele Treffer gibt, wird nur ein Teil davon angezeigt, d.h. befindet sich der gesuchte Eintrag nicht in der Liste, müssen Sie evtl. weitere Buchstaben dazu eingeben.

Durch Anklicken eines Eintrags in der Liste wird dieser in das Eingabefeld übernommen (alternativ: Auswählen mit den Pfeiltasten und Drücken der „Tabulator“-Taste). Das Textfeld akzeptiert aber auch beliebige andere Eingaben; es muss nicht zwingend ein Eintrag aus der Liste übernommen werden!

Die **Checkbox für „bestätigte“ Lemmata** wird automatisch markiert, wenn ein grün hervorgehobener Eintrag aus der Liste ausgewählt wird. Ebenso wird eine Markierung automatisch wieder entfernt, wenn der Lemma-Eintrag auf andere Weise geändert wird. Dies soll vor unabsichtlichen Fehlern (und damit vor falschen Einträgen, die trotzdem „bestätigt“ sind) schützen.

Klicken auf das Symbol  öffnet die Online-Version des Grimmschen Wörterbuchs in einem separaten Browserfenster. Dabei wird automatisch der zu dem eingetragenen Lemma passende Eintrag aufgerufen. Weitere Aufrufe öffnen sich immer in demselben Fenster.



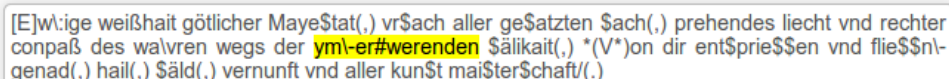



Abbildung 4: Horizontale Textansicht

### 3.2.7 Kommentarfeld

Die Spalte „Kommentar“ enthält ein Eingabefeld, worin ein beliebiger Freitext eingegeben werden kann. Diese Kommentare sind unabhängig von Kommentaren in der Transkription (+K ... @K).

### 3.2.8 Kontextmenü

Klicken auf den Pfeil  öffnet ein Kontextmenü. Derzeit besteht es nur aus drei Einträgen, die sich alle auf die Bearbeitung der Transkription beziehen, welcher der Abschnitt 4 gewidmet ist.

## 3.3 Horizontale Textansicht

Am unteren Rand des Editors befindet sich eine Textansicht, die einen Ausschnitt aus dem derzeit geöffneten Dokument wiedergibt. Dort können die Token nebeneinander (horizontal) statt untereinander gelesen werden, um den Satzzusammenhang schneller zu erfassen. Das Token, das gerade annotiert wird, ist in dieser Ansicht gelb hinterlegt (s. Abbildung 4). Diese Markierung ändert sich immer dann, wenn ein Textfeld oder eine Dropdown-Box im Editor angewählt bzw. bearbeitet wird.

Die horizontale Textansicht umfasst immer mindestens die Token, die auf der aktuell angezeigten Seite sichtbar sind. Nach Möglichkeit erweitert CorA die Anzeige immer bis zu einer Satzgrenze, damit der gesamte Satzzusammenhang gesehen werden kann. Es gibt jedoch derzeit ein Limit nach oben, wieviele Token als weiterer „Kontext“ angezeigt werden.

Aussehen und Inhalt der horizontalen Textansicht lassen sich z. Zt. noch nicht näher konfigurieren; dies ist eventuell zu einem späteren Zeitpunkt möglich.

## 3.4 Anpassen des Editors

Das Aussehen des Editors kann in verschiedenen Punkten individuell angepasst werden.

Die **Anordnung der Spalten** kann per „Drag & Drop“ geändert werden. Gehen Sie dazu auf eine beliebige Spalten-Überschrift, klicken und halten Sie die linke Maustaste gedrückt, ziehen Sie die Spalte an eine beliebige andere Position, und lassen die Maustaste wieder los. CorA merkt sich die individuelle Anordnung der Spalten automatisch.

Weitere Anpassungen können **über den Reiter „Einstellungen“** vorgenommen werden (s. Abbildung 5). Die Einstellung **Zeilen pro Seite** steuert, wieviele Wortformen im Editor gleichzeitig dargestellt werden. Wir empfehlen, aus Performance-Gründen nicht mehr als 50 Zeilen pro Seite anzeigen zu lassen. **Überlappende Zeilen** gibt an, wieviele Zeilen vom Seitenende auch noch auf

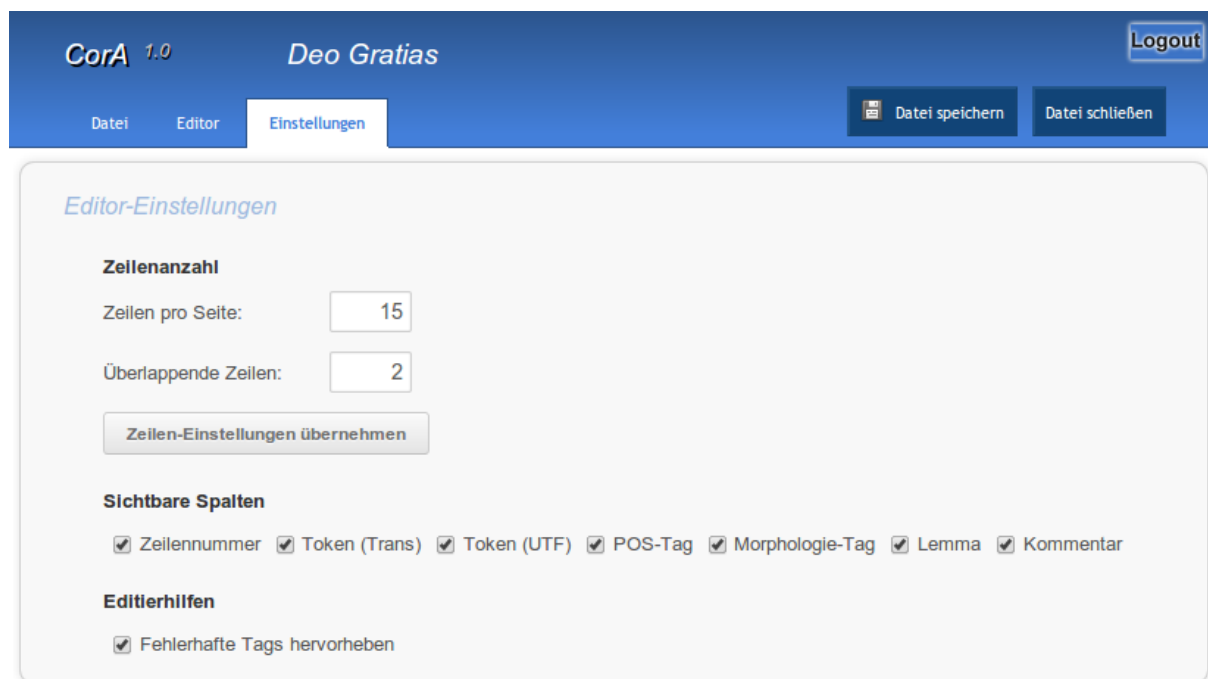


Abbildung 5: Der Reiter „Einstellungen“

dem Anfang der nächsten Seite erscheinen sollen. Beide Einstellungen müssen durch Klick auf „Zeilen-Einstellungen übernehmen“ bestätigt werden.

Die Checkboxen unter **Sichtbare Spalten** steuern, welche Spalten im Editor angezeigt werden. Standardmäßig sind alle Spalten sichtbar, durch Abwählen einzelner Checkboxen können bestimmte Spalten jedoch ausgeblendet werden. Dies kann der besseren Übersichtlichkeit im Editor dienen. Änderungen hier müssen nicht separat bestätigt werden und sind sofort wirksam.

**Hinweis**

Wenn Sie Spalten sichtbar machen, die zuvor ausgeblendet waren, kann es vereinzelt zu Darstellungsfehlern im Editor kommen. Neuladen der Seite sollte dieses Problem beheben. (Speichern Sie eventuelle Änderungen vorher ab!)


Die Option **Fehlerhafte Tags hervorheben** steuert, ob fehlerhafte Tags im Editor mit einer roten Umrandung versehen werden.

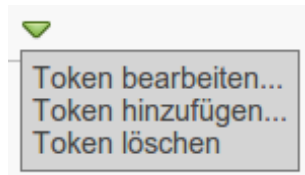
## 4 Bearbeiten der Transkription

CorA ist in erster Linie ein Annotationstool, stellt jedoch auch Funktionen für den Fall bereit, dass Änderungen an der Original-Transkription vorgenommen werden müssen. Die dafür nötige Vorgehensweise kann anfangs – je nach Art der Änderung – möglicherweise etwas unintuitiv sein. Dies ist zumeist dann der Fall, wenn die Änderungen sich auf die Tokenisierung auswirken (z.B. wenn zwei Wortformen, die in CorA auf zwei Zeilen verteilt sind, zusammengezogen werden sollen). Im Folgenden werden die einzelnen Funktionen beschrieben und an einigen praktischen Beispielen erläutert.

### Wichtig!

- Vor jeder Änderung der Transkription müssen eventuell geänderte Annotationen **gespeichert** werden! Ungespeicherte Änderungen am Dokument gehen ansonsten verloren. (CorA weist Sie auch ggf. mit einem roten Hinweistext darauf hin.)
- Änderungen an der Transkription sind augenblicklich gespeichert und können **nicht rückgängig** gemacht werden! Sie können die Transkription zwar per Hand wieder zurück ändern, es gibt jedoch keine Möglichkeit, den Zustand der Transkription vor der Bearbeitung zu sehen.

Durch Klick auf das Symbol  (standardmäßig ganz am Ende jeder Zeile) öffnet sich ein Kontextmenü, worüber alle Funktionen zum Ändern der Transkription aufgerufen werden können:



Die Funktion „Token bearbeiten“ kann außerdem aufgerufen werden durch **Doppelklick** auf die entsprechende Wortform in der Spalte „Token (Trans)“ oder „Token (UTF)“.

### 4.1 Token bearbeiten

Für die Bearbeitung der Transkription ist es wichtig zu verstehen, dass die Zeilen im Editor und die Transkription, die bearbeitet wird, sich nicht immer 1:1 entsprechen. Dies lässt sich am besten an einem Beispiel erläutern: Abbildung 6a zeigt einen Ausschnitt von drei Zeilen aus dem Editor; Abbildung 6b zeigt das Fenster, das sich öffnet, wenn eine beliebige dieser drei Zeilen bearbeitet wird. Die Transkription „maister\$chaft/(.)“ wird für die Annotation in die drei separaten Token „maister\$chaft“, „/“ und „(.)“ zerlegt, während die Bearbeitung dieser Transkription jedoch als Ganzes erfolgt.

Im Editorfenster aus Abbildung 6b können nun beliebige Änderungen vorgenommen werden, die nach einem Klick auf „Speichern“ übernommen werden. Für alle Änderungen gilt, dass diese zunächst mit Hilfe des Check-Skripts geprüft werden. Ungültige Transkriptionszeichen werden von CorA

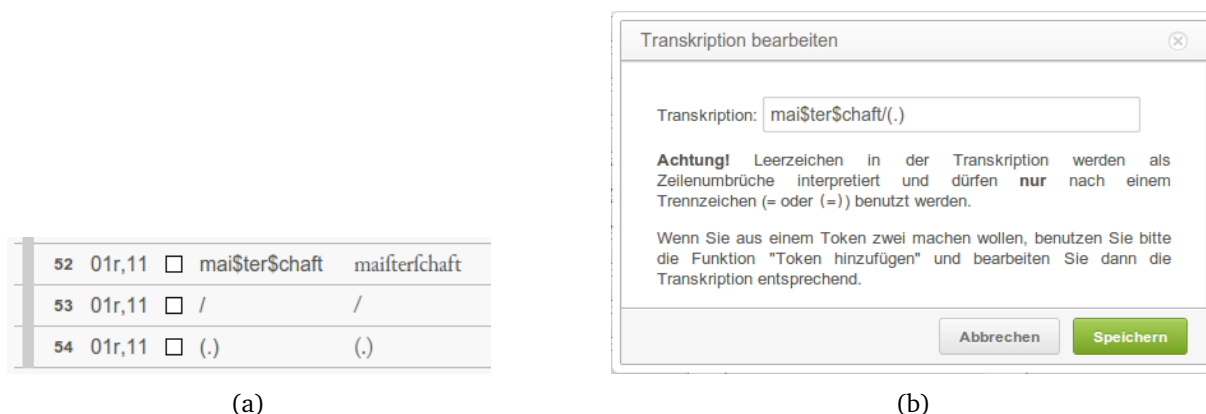


Abbildung 6: Bearbeiten der Original-Transkription

nicht akzeptiert; die entsprechende Fehlermeldung des Check-Skripts wird dann angezeigt und die Änderung wird nicht vorgenommen.

**Leerzeichen in der Transkription** haben an dieser Stelle eine besondere Bedeutung. Sie werden von CorA immer dann verwendet, wenn sich ein Token in der Original-Transkription über mehrere Zeilen erstreckt:

```
F137-02r,20    dein hohe vernunft $o begirlich be=
F137-02r,21    gert(,) $ucht vnd erfra\vgt(,) alle kun$t
```

Das Token, das sich hier über beide Zeilen erstreckt, würde beim Bearbeiten in CorA als „be= gert(,)“ angezeigt. Das Leerzeichen markiert dabei die Stelle, an der der Zeilenumbruch stattfindet. Dieses Leerzeichen darf beim Bearbeiten der Transkription nicht entfernt werden, da ansonsten der hintere Teil in der Original-Transkription mit auf die vorherige Zeile gezogen würde!

Andersherum dürfen Leerzeichen beim Bearbeiten der Transkription nur dann eingefügt werden, wenn Sie einen Zeilenumbruch repräsentieren sollen. Dies ist nur möglich, wenn das bearbeitete Token am Ende einer Zeile steht und wird ansonsten von CorA nicht akzeptiert. Außerdem muss jedem Leerzeichen ein Trennzeichen vorausgehen.

#### Tipp

Möchten Sie einmal in die Transkription tatsächlich ein Leerzeichen einfügen, das nicht für einen Zeilenumbruch steht (z.B. um „in#der“ in „in der“ abzuändern), so sind dafür mehrere Schritte nötig! In Abschnitt 4.4 wird dies an Beispielen verdeutlicht.

## 4.2 Token hinzufügen

Wählen Sie „Token hinzufügen...“ aus dem Kontextmenü aus, um eine neue Transkription **vor** einem bestehenden Token einzufügen. Es öffnet sich ein Fenster, worin die neue Transkription

Transkription hinzufügen

Transkription:

Die neue Transkription wird **vor** dem Token "maister\$chaft/(.)" auf Zeile "01r,11" in das Originaldokument eingefügt.

Abbrechen Speichern

Abbildung 7: Hinzufügen eines Tokens

eingetragen werden kann (s. Abbildung 7). Dabei gelten dieselben Regeln wie für das Bearbeiten einer Transkription (siehe oben); insbesondere darf die Transkription kein Leerzeichen enthalten.

In einem Hinweistext unter dem Eingabefeld wird nochmals erläutert, an welcher Position und auf welcher Zeile der Original-Transkription das neue Token eingefügt wird. Überprüfen Sie diese Angaben in jedem Fall, um sicherzustellen, dass Sie das Token an der richtigen Stelle im Text einfügen!

### 4.3 Token löschen

Wählen Sie „Token löschen“ aus dem Kontextmenü aus, um eine Transkription komplett aus dem Text zu löschen. Sie müssen diese Aktion zur Sicherheit immer bestätigen, da auch das Löschen eines Tokens nicht rückgängig gemacht werden kann. In dem entsprechenden Hinweistext können Sie außerdem nochmals überprüfen, welche Transkription bei einem Bestätigen der Aktion tatsächlich gelöscht wird.

### 4.4 Praktische Beispiele

Für Änderungen in der Transkription, die Einfluss auf die Tokenisierung haben, können mitunter mehrere Bearbeitungsschritte notwendig sein. Daher werden hier einige Beispiele gegeben, wie die Transkription in diesen Fällen geändert werden kann.

**Beispiel 1.** Angenommen, ein Token in der Transkription soll (durch Einfügen eines Leerzeichens) in zwei Token aufgetrennt werden. Ein Beispiel wäre die Änderung von „in#der“ zu „in der“. Hierzu sind zwei Schritte nötig:

1. Gehen Sie auf die Zeile mit der Transkription „in#der“, öffnen Sie das Kontextmenü (👉), und wählen Sie „Token hinzufügen“. Fügen Sie jetzt „in“ als neue Transkription ein.  
→ In der Transkription steht jetzt „in in#der“.
2. Doppelklicken Sie nun auf die Transkription „in#der“ und ändern Sie sie ab zu „der“.  
→ In der Transkription steht jetzt „in der“.

**Beispiel 2.** Angenommen, zwei Token in der Transkription sollen zu einem zusammengefügt werden. Beispielsweise soll „zu vor“ geändert werden in „zu#vor“. Dies ist der umgekehrte Fall zu Beispiel 1. Hierfür sind ebenfalls wieder zwei Schritte nötig:

1. Doppelklicken Sie auf die Transkription „vor“ und ändern Sie sie ab zu „zu#vor“.  
→ In der Transkription steht jetzt „zu zu#vor“.
2. Gehen Sie nun auf die Zeile mit der Transkription „zu“, öffnen Sie das Kontextmenü, und wählen Sie „Token löschen“.  
→ In der Transkription steht jetzt „zu#vor“.

### Tipp

Seien Sie besonders vorsichtig, wenn sie eine Transkription bearbeiten, die sich über mehrere Zeilen erstreckt! Angenommen, in der folgenden Original-Transkription sollen die ersten beiden Token zu „MVLLNER#ORD= nung“ zusammengefügt werden:

```
F57-1r,01    MVLLNER ORD=
F57-1r,02    nung in den Für$tlichen Thiroli$chen Stet=
```

Dies kann wie oben beschrieben erreicht werden, indem Sie in CorA zunächst das Token „ORD= nung“ abändern in „MVLLNER#ORD= nung“, und anschließend das erste Token „MVLLNER“ löschen.

Andersherum ist es jedoch **nicht** möglich, zuerst „MVLLNER“ in „MVLLNER#ORD= nung“ zu ändern und dann das Folgetoken zu löschen! Die entsprechende Operation wird fehlschlagen, da die neue Transkription ein Leerzeichen (= Zeilenumbruch) enthält, das Token „MVLLNER“ sich jedoch nicht am Zeilenende befindet.

Behelfen Sie sich **in keinem Fall**, indem Sie das Leerzeichen weglassen oder entfernen! Das Ergebnis würde dann fälschlicherweise so interpretiert...

```
F57-1r,01    MVLLNER#ORD=nung
F57-1r,02    in den Für$tlichen Thiroli$chen Stet=
```

... was vermutlich nicht Ihrer Absicht entspricht.

**Beispiel 3.** Angenommen, ein fälschlich gesetztes Trennzeichen am Zeilenende soll entfernt werden. Beispielsweise soll „in(=) der“ zu „in der“ geändert werden, wobei die beiden Wortformen auf verschiedenen Zeilen in der Transkription stehen. Dies kann mit folgenden Schritten realisiert werden:

1. Gehen Sie auf die erste Zeile **nach** der Transkription „in(=) der“, öffnen Sie das Kontextmenü, und wählen Sie „Token hinzufügen“. Fügen Sie jetzt „der“ als neue Transkription ein.  
→ In der Transkription steht jetzt „in(=) der der“.
2. Doppelklicken Sie nun auf die Transkription „in(=) der“ und ändern Sie sie zu „in“.  
→ In der Transkription steht jetzt „in der“.



**Beispiel 4.** Angenommen, das Token „ge“ am Ende einer Zeile soll mit dem ersten Token „schah“ der nächsten Zeile mittels eines Trennstriches „(=)“ zusammengeführt werden. Dies gelingt mit folgenden Schritten:

1. Doppelklicken Sie auf die Transkription „ge“ und ändern Sie sie zu „ge(=) schah“ (Leerzeichen beachten!).
2. Löschen Sie nun einfach das zweite „schah“.

## 5 Tastaturbefehle

Im Folgenden werden einige Tastaturbefehle aufgeführt, die für ein zügigeres Arbeiten in CorA hilfreich sein können. Diese Liste wird im Laufe der Zeit noch ergänzt.

Mit **Tab** (Tabulator-Taste) springen Sie im Editor zum nächsten editierbaren Feld, mit **Shift+Tab** springen Sie ein Feld zurück. (Dies ist eine Funktion Ihres Internetbrowsers, nicht von CorA selbst.)

In den Dropdown-Menüs für POS- und Morphologie-Tags können Sie mit **Pfeiltaste nach oben/unten** zwischen den Tags wechseln. Dasselbe gilt für das Menü mit den Lemma-Vorschlägen. Alternativ können Sie auch die Anfangsbuchstaben des gewünschten Tags eingeben. (Dies ist eine Funktion Ihres Internetbrowsers, nicht von CorA selbst.)

Mit **Strg+S** können Sie die aktuelle Datei speichern.

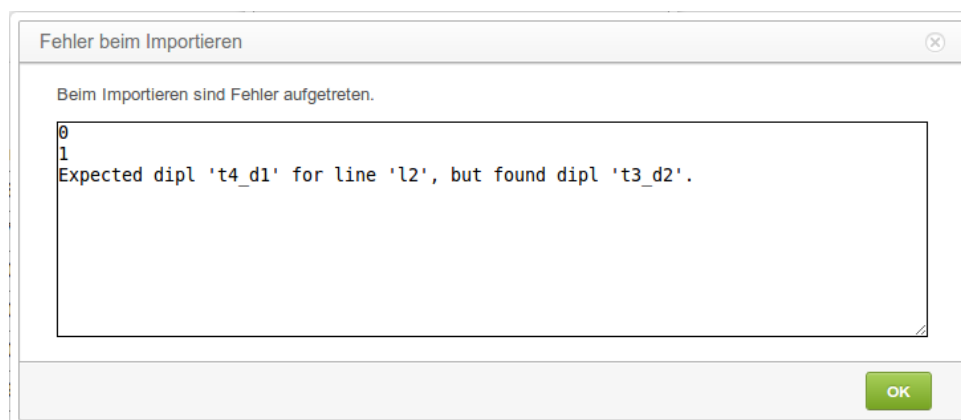


Abbildung 8: Eine Fehlermeldung beim Import

## 6 Umgang mit Fehlermeldungen

Trotz aller Bemühungen können wir leider nicht ganz ausschließen, dass während der Benutzung von CorA auch Fehlermeldungen auftreten. In unvorhergesehenen Fällen können diese Meldungen mitunter recht kryptisch sein; Abbildung 8 zeigt ein Beispiel für eine (hoffentlich niemals auftretende...) Fehlermeldung beim Importieren einer Datei. Damit wir in so einem Fall das zugrundeliegende Problem so schnell wie möglich beheben können, bitten wir Sie darum, ein paar einfache Hinweise zu beachten.

1. Melden Sie Fehler immer per **E-Mail** an Marcel Bollmann ([bollmann@linguistics.rub.de](mailto:bollmann@linguistics.rub.de)).
2. Melden Sie uns Fehler, die Sie sich nicht erklären können, bitte in jedem Fall! Wir können nur Fehler beheben, von denen wir auch wissen.
3. Senden Sie uns eine **aussagekräftige** Beschreibung des Fehlers! Dazu sollten immer folgende Informationen gehören:
  - Was genau haben Sie unmittelbar vor dem Auftreten des Fehlers getan? (z.B. *auf den Knopf „Datei speichern“ geklickt*)
  - Wie genau äußert sich der Fehler? Falls es eine Fehlermeldung gibt: bitte die komplette Meldung kopieren und einfügen!
  - An welcher Datei oder welchem Token genau haben Sie gearbeitet, als der Fehler auftrat? Falls der Fehler beim Importieren auftritt: die Datei, die importiert werden sollte, als Anhang mitschicken!