

CorA (Corpus Annotator) Benutzerhandbuch

Version 1.2

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

27. Januar 2015

Inhaltsverzeichnis

1 Allgemeines	3
1.1 Benutzername und Passwort	3
1.2 Die Titelleiste	3
2 Dateiverwaltung	5
2.1 Importieren von Transkriptionen	6
3 Das Editor-Fenster	7
3.1 Annotationsebenen (Spalten im Editor)	7
3.1.1 Fortschrittsbalken (P)	7
3.1.2 Nummerierung	8
3.1.3 Markierung von Problemfällen (E)	8
3.1.4 Token	8
3.1.5 POS- und Morphologie-Tags	8
3.1.6 Lemmatisierung	9
3.1.7 Normalisierung & Modernisierung	10
3.1.8 Kommentarfeld	10
3.1.9 Kontextmenü	10
3.2 Die Toolbar	11
3.3 Suchen innerhalb des Dokuments	12
3.4 Automatische Annotation mit externen Tools	14
3.5 Anpassen des Editors	15
4 Bearbeiten der Transkription	16
4.1 Token bearbeiten	17
4.2 Token hinzufügen	17
4.3 Token löschen	17

4.4	Praktische Beispiele	18
5	Tastaturbefehle	19
6	Umgang mit Fehlermeldungen	20

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 Chrome¹ (bzw. Chromium) oder Mozilla Firefox². 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).

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!

1.2 Die Titelleiste

In der Titelleiste, die sich immer am oberen Bildschirmrand befindet, können Sie

- Sigle und Dateinamen des aktuell geöffneten Dokuments sehen;
- Benutzernamen und Online-Status sehen (das Symbol in der oberen rechten Ecke ist **grün**, wenn alles okay ist, und wird **rot**, falls die Verbindung zum Server verloren geht bzw. etwaige Fehler auftreten);
- zwischen den verschiedenen Tabs wechseln (s.u.);
- die aktuell geöffnete **Datei schließen** bzw. sich aus CorA **ausloggen**.

¹<http://www.google.com/chrome/>

²<http://www.mozilla.com/>

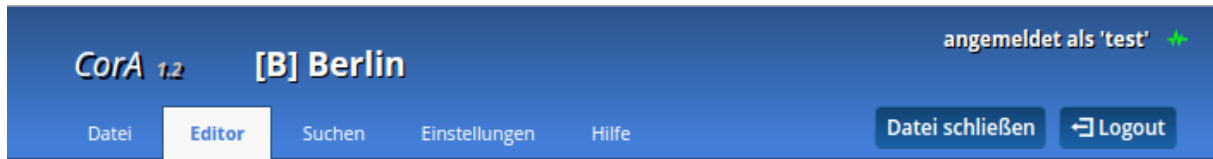


Abbildung 1: Titelleiste von CorA

Hier eine Übersicht über die verschiedenen Tabs:

- **Datei:** Hier können Dateien geöffnet, importiert, exportiert, oder gelöscht werden (s. Abschnitt 2).
- **Editor:** Das Haupt-Editorfenster (s. Abschnitt 3); nur sichtbar, wenn aktuell eine Datei geöffnet ist.
- **Suchen:** Hier erscheinen die Ergebnisse einer Suchanfrage (s. Abschnitt 3.3); nur sichtbar, nachdem eine Suchanfrage gestartet wurde.
- **Einstellungen:** Hier können Anpassungen am Editor vorgenommen werden (s. Abschnitt 3.5).
- **Hilfe:** Hier findet sich u.a. immer die aktuellste Version dieses Benutzerhandbuchs.

2 Dateiverwaltung

Nach der Anmeldung in CorA wird zunächst der Reiter „Datei“ angezeigt:

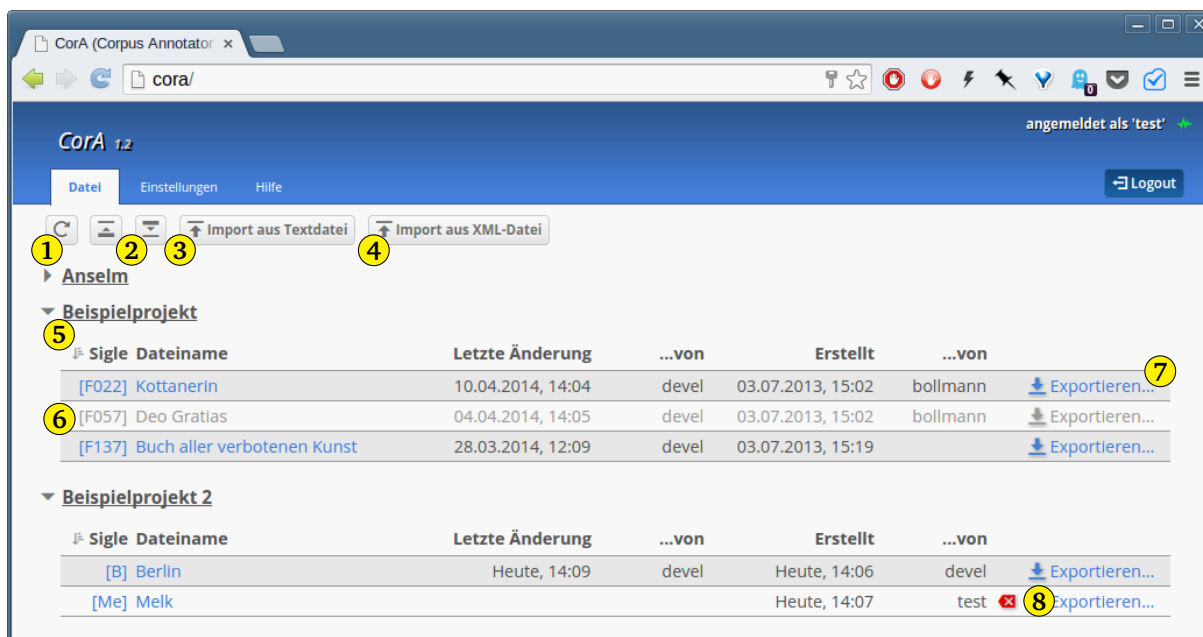


Abbildung 2: Dateien verwalten

Alle Dateien in CorA sind **Projekten** zugeordnet (hier z.B. “Beispielprojekt”). Der Button ① aktualisiert die Dateiliste. Ein Klick auf den Projektnamen klappt die Liste der zugeordneten Dateien ein bzw. aus. Die Buttons ② klappen die Dateilisten aller Projekte gleichzeitig ein bzw. aus.

Neue Texte importieren können Sie entweder als Textdatei ③ oder als CorA-XML-Datei ④. Der Import als Textdatei ist nur möglich, wenn dem entsprechenden Projekt ein Importskript zugeordnet wurde, der das Textformat intern nach CorA-XML konvertiert. (Details zum Import finden sich in Abschnitt 2.1.)

Die Dateien sind standardmäßig aufsteigend nach Sigle **sortiert** ⑤. Durch Klick auf eine beliebige Spaltenüberschrift können Sie die Dateiliste nach dieser Spalte sortieren lassen, um z.B. schnell alle kürzlich bearbeiteten Dateien zu sehen. Klicken Sie mehrmals auf dieselbe Spalte, um zwischen aufsteigender und absteigender Sortierung zu wechseln.

Um eine Datei zu öffnen, klicken Sie einfach auf einen Dateinamen. Dateien, die **leicht ausgegraut** erscheinen ⑥, 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 die Datei schließt, sich ganz aus CorA ausloggt, oder länger als 30 Minuten nicht mehr online gewesen ist.

“**Exportieren...**” ⑦ öffnet ein Fenster, um eine Datei im CorA-XML-Format oder als mehrspaltige Textdatei zu exportieren. Ein Klick auf ⑧ **löscht** eine Datei komplett; dieser Vorgang ist endgültig und **kann nicht rückgängig gemacht werden!** Nur Administratoren oder der Benutzer, der die entsprechende Datei importiert hat, kann sie auch wieder löschen.

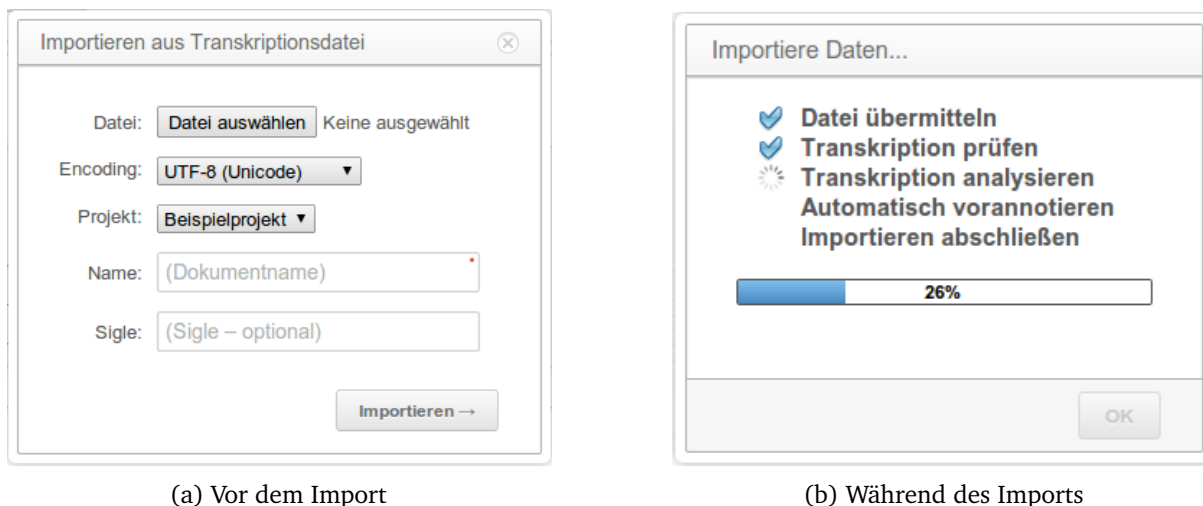


Abbildung 3: Importieren einer Datei

2.1 Importieren von Transkriptionen

Neue Texte können jederzeit im CorA-XML-Format importiert werden. Wenn für ein Projekt ein Importskript festgelegt ist,³ können auch Textdateien, z.B. Transkriptionsdateien, direkt importiert werden via „Import aus Textdatei“ (vgl. ③ in Abbildung 2).

Es öffnet sich ein Fenster ähnlich wie in Abbildung 3a. 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 3b). 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!⁴ 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!

³Dies können nur Administratoren einrichten.

⁴Wenn 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.

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.

Änderungen, die Sie im Editor vornehmen, werden von CorA **automatisch gespeichert**. CorA wird Sie ggf. darüber informieren, wenn der Speichervorgang fehlschlägt.

Wenn Sie sichergehen wollen, dass Ihre Änderungen auch wirklich gespeichert wurden, klicken Sie immer explizit auf **“Datei schließen”** oder **“Logout”**, bevor Sie das Browserfenster schließen! Wenn die Datei ordnungsgemäß geschlossen bzw. Sie ordnungsgemäß ausgeloggt wurden, sind Ihre gemachten Änderungen auf jeden Fall übernommen worden.

Falls es ein Problem mit der Internetverbindung oder einen sonstigen Fehler beim Speichervorgang gibt, wird CorA Sie spätestens beim Versuch, die Datei zu schließen bzw. sich auszuloggen ausdrücklich darauf hinweisen, und den entsprechenden Vorgang abbrechen!

3.1 Annotationsebenen (Spalten im Editor)

Wortformen im CorA-Editor werden immer zeilenweise angezeigt, wobei jede Zeile ein Token nach moderner Tokenisierung enthält. Die Spalten in der Editor-Tabelle entsprechen verschiedenen Annotationsebenen, wobei die verfügbaren Annotationsebenen flexibel pro Text eingestellt werden können.⁵ Nicht jede der im folgenden aufgeführten Ebenen ist daher für jeden Text verfügbar!

3.1.1 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, und überschreibt nur Daten, die noch nicht grün markiert sind.

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.

⁵Dies kann jedoch nur ein Administrator festlegen.

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.

3.1.2 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

3.1.3 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 kann z.B. benutzt werden, um Fehler oder Problemfälle zu markieren, die später noch einmal angesehen werden müssen. Derart markierte Tokens lassen sich später z.B. mithilfe der Suchfunktion gezielt auffinden.

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

3.1.4 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 3.5.

3.1.5 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.

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.

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

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).

3.1.6 Lemmatisierung


Die Lemma-Spalte hat drei Bestandteile: eine **Checkbox**, um Lemma-Einträge als „bestätigt“ zu markieren; ein **Eingabefeld** für das Lemma; und ggf. einen **Verweis** auf eine externe Resource (z.B. Online-Ausgabe eines Lexikons).

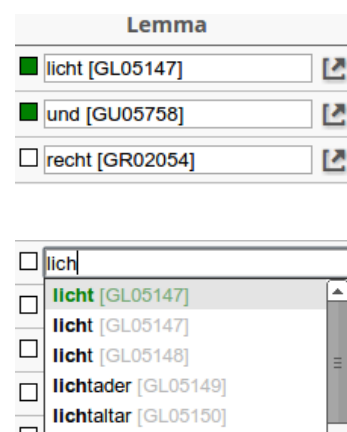
Das Eingabefeld ist mit einer **Dropdown-Liste mit Lemma-Vorschlägen** verknüpft. Diese Liste öffnet sich automatisch, sobald eine Eingabe in dem Feld gemacht wird (s. Abbildung rechts). Alternativ lässt sie sich auch durch Doppelklick in das Eingabefeld oder mit „Pfeiltaste nach unten“ öffnen. Die Liste 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.
3. Alle anderen Einträge sind durch Autovervollständigung gefundene Treffer in einer verlinkten Lemma-Liste (nur, falls eine solche Liste für das Projekt definiert worden ist). Dabei gilt:
 - a) es wird immer nur nach passendem Wortanfang gesucht;
 - b) Diakritika 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 eine evtl. verlinkte externe Resource in einem separaten Browserfenster. Dabei wird automatisch der zu dem eingetragenen Lemma passende Eintrag aufgerufen. Weitere Aufrufe öffnen sich immer in demselben Fenster.



3.1.7 Normalisierung & Modernisierung

Die Spalten “Normalisierung” und “Modernisierung” sind einfache Textfelder, die beliebige Eingaben akzeptieren. Falls die Spalte “Modernisierung” leer ist, erscheint standardmäßig der Wert der Normalisierung grau hinterlegt. Dies spiegelt die Konvention wider, dass die Modernisierung nur angegeben werden muss, wenn sie sich von der Normalisierung unterscheidet.


Normalisierung	Modernisierung	
nun	nun	▼
helfe	helfen	f ▼
sie	sie	▼
uns	uns	▼

In der Dropdown-Box neben dem Modernisierungsfeld kann der Typ der Modernisierung (z.B. “f” = “durch Flexion bedingt”, “s” = “durch Semantik bedingt”) näher spezifiziert werden. Ist die Modernisierung leer, bleibt diese Dropdown-Box deaktiviert; wird in das Modernisierungsfeld etwas eingetragen, so muss auch ein Wert in der Dropdown-Box ausgewählt werden. Andernfalls wird sie rot umrandet dargestellt, um auf die fehlende Annotation hinzuweisen.

3.1.8 Kommentarfeld

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

3.1.9 Kontextmenü

Klicken auf den Pfeil  öffnet ein Kontextmenü. Derzeit besteht es aus folgenden Einträgen:

- **“Suche ähnliche...”** öffnet den Suchdialog mit voreingestellten Werten (vgl. Abs. 3.3);
- **“Token bearbeiten...”**, **“Token hinzufügen...”** und **“Token löschen”** ermöglichen die Bearbeitung der Original-Transkription, welcher der Abschnitt 4 gewidmet ist.

3.2 Die Toolbar

Am oberen Rand des Editors befindet sich eine Toolbar, die zur Navigation innerhalb des Dokuments und zum Aufrufen verschiedener Funktionen verwendet werden kann.



Abbildung 4: Die Editor-Toolbar

- ① Zeigt die **aktuelle Seitenzahl** und die Gesamtanzahl an Seiten an. Durch Klick auf dieses Element kann direkt zu einer bestimmten Seite im Editor navigiert werden.
- ② **Blättert** eine Seite vor bzw. zurück.
- ③ Springt zu einer bestimmten Zeile innerhalb des Dokuments.
- ④ Macht die letzte Änderung **rückgängig** bzw. **stellt sie wieder her**.
- ⑤ Aktiviert die **Suchfunktion** (vgl. Abs. 3.3).
- ⑥ Springt zum vorherigen bzw. nächsten Suchergebnis. Nur aktiv, wenn zuvor eine Suche ausgeführt wurde.
- ⑦ Startet die **automatische Annotation** mithilfe externer Tools (vgl. Abs. 3.4).
- ⑧ Zeigt die **Metadaten** des aktuellen Dokuments an bzw. erlaubt, sie zu ändern.

Zeilen- und Seitenzahlen beziehen sich dabei immer auf die Darstellung im Editor, **nicht** auf Zeilen bzw. Seiten in der Transkription! Die Seitenzahlen sind abhängig davon, wieviele Zeilen pro Seite angezeigt werden, was vom Benutzer selbst eingestellt werden kann (vgl. Abschnitt 3.5). Daher können sie auch für denselben Text von Benutzer zu Benutzer verschieden sein!

Funktionen zum Navigieren anhand der Nummerierung in der Original-Transkription sind momentan noch nicht implementiert.

3.3 Suchen innerhalb des Dokuments

Durch Klick auf “Suchen...” im Editor oder “Suche ändern” in den Suchergebnissen kann eine neue Suche gestartet werden. CorA kann nur im aktuell geöffneten Dokument suchen, und nur nach einzelnen Wortformen – eine kontextabhängige Suchanfrage (z.B. “Token x gefolgt von y”) ist nicht möglich.

Abbildung 5: Suchfenster

Folgende Einstellungen können bei der Suche vorgenommen werden:

- ① Steuert, ob **alle** oder nur **mindestens eins** der angegebenen Suchkriterien erfüllt sein müssen.
- ② Definiert die Suchkriterien, d.h. welches **Feld** (z.B. Token, POS-Tag) auf welche Weise (z.B. “ist”, “endet auf”) mit welchem **Wert** verglichen werden soll.
- ③ Entfernt ein Suchkriterium bzw. fügt ein neues hinzu.
- ④ Löscht alle momentan angegebenen Suchkriterien.

Die Suchergebnisse werden in einem separaten Tab “Suchen” angezeigt. Annotationen können hier genauso bearbeitet werden wie im Haupteditor auch:

P	#	Zeile	E	Token (Trans)	Token (UTF)	POS-Tag	Morphologie-Tag	Kommentar
	395	3r,17	<input type="checkbox"/>	wer\$t	werft	VVFIN	***	
	530	4r,07	<input type="checkbox"/>	i\$t	ift	VVFIN	***	
	543	4r,10	<input type="checkbox"/>	i\$t	ift	VVFIN	***	
	647	4v,11	<input type="checkbox"/>	i\$t	ift	VVFIN	***	
	812	5v,06	<input type="checkbox"/>	vorre\$t	vorreft	VVFIN	***	

Abbildung 6: Anzeige der Suchergebnisse

Ein Klick auf die Zeilennummer oder das Token **springt zur entsprechenden Textstelle im Editor**. Sie können durch Klick auf den jeweiligen Tab jederzeit zwischen Editor und Suchergebnissen hin und

her wechseln. Zum Navigieren innerhalb der Suchergebnisse können außerdem die entsprechenden Buttons im Editor (6 in Abb. 4) verwendet werden.

Um eine neue Suche zu starten, klicken Sie einfach wieder auf “Suchen...” bzw. “Suche ändern”. Ihre letzten Suchkriterien bleiben erhalten, bis Sie im Suchfenster auf “Zurücksetzen” klicken, sodass Sie z.B. auch ihre vorherige Suche modifizieren oder erneut ausführen können.

Beachten Sie, dass die Suchergebnisse sich **nicht** automatisch aktualisieren, wenn Sie zwischenzeitlich Änderungen im Editor vornehmen. Wenn Sie also z.B. nach Tokens mit dem POS-Tag XX suchen, und anschließend bei einigen Tokens den POS-Tag XX auf YY ändern, werden diese dennoch weiterhin in den Suchergebnissen angezeigt. Sie können jedoch einfach das Suchfenster öffnen und die Suche mit denselben Kriterien erneut starten, um die Ergebnisse zu aktualisieren.

Um nach Tokens zu suchen, die einem bestimmten Token in der Editor-Ansicht ähnlich sind, gibt es im **Dropdown-Menü** des jeweiligen Tokens den Eintrag “**Suche ähnliche...**”. Dieser Menüpunkt öffnet das Suchfenster, wobei die Suchkriterien automatisch so ausgefüllt sind, dass sie den Annotationen des aktuell ausgewählten Tokens entsprechen. Die Suche kann entweder direkt ausgeführt oder noch weiter modifiziert werden.

3.4 Automatische Annotation mit externen Tools

CorA bietet die Möglichkeit, direkt aus der Weboberfläche heraus externe Annotationstools aufzurufen. Hierzu klicken Sie in der Editor-Ansicht auf den Knopf “Automatisch annotieren” (vgl. 7 in Abb. 4); es sollte sich ein Fenster ähnlich wie dieses öffnen:



Abbildung 7: Starten der automatischen Annotation

Hier können Sie zwischen verschiedenen Taggern auswählen, die auf das aktuell geöffnete Dokument anwendbar sind. Die Einrichtung dieser Tagger kann nur ein Administrator vornehmen; bitte fragen Sie ggf. dort nach, wenn Sie sich nicht sicher sind, was ein bestimmter Tagger in der Auswahlliste genau macht, oder falls ein neuer Tagger hinzugefügt werden soll.

Durch Klick auf “**Annotieren**” wird der Tagger aufgerufen und auf das aktuell geöffnete Dokument angewendet; dabei werden alle eventuell bereits bestehenden Annotationen **überschrieben**, die **nicht** mittels des Fortschrittsbalkens grün markiert sind.

Das automatische Annotieren kann **nicht rückgängig** gemacht werden! Bitte stellen Sie vor dem Klick auf “Annotieren” sicher, dass der grüne Fortschrittsbalken an der korrekten Position steht.

Der Button “**Neu trainieren**” ist je nach Konfiguration nicht für jeden Tagger verfügbar. Falls er aktiv ist, wird hierdurch der ausgewählte Tagger auf dem geöffneten Text **und allen Texten des übergeordneten Projekts** neu trainiert. Hierfür werden **nur** Textpassagen herangezogen, die mittels des Fortschrittsbalkens grün markiert sind.



Abbildung 8: Der Reiter „Einstellungen“

3.5 Anpassen des Editors

Der Editor kann über den **Reiter „Einstellungen“** individuell angepasst werden (s. Abbildung 8).

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

“Sichtbare Spalten” steuert, welche Spalten im Editor angezeigt werden. Standardmäßig sind alle Spalten sichtbar, durch Abwählen einzelner Checkboxes können bestimmte Spalten jedoch ausgeblendet werden. Dies kann der besseren Übersichtlichkeit im Editor dienen. Änderungen hier sind sofort wirksam.

Unter **“Horizontale Textvorschau”** kann eingestellt werden, welche Tokenebene (Trans oder UTF) in der Textvorschau am unteren Bildschirmrand angezeigt werden soll. Alternativ kann die Textvorschau hier auch deaktiviert werden.

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




Abbildung 9: Bearbeiten der Original-Transkription

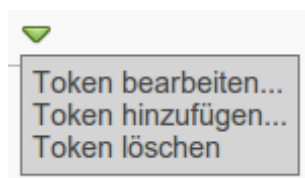
4 Bearbeiten der Transkription

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!

- Änderungen an der Transkription 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)“.

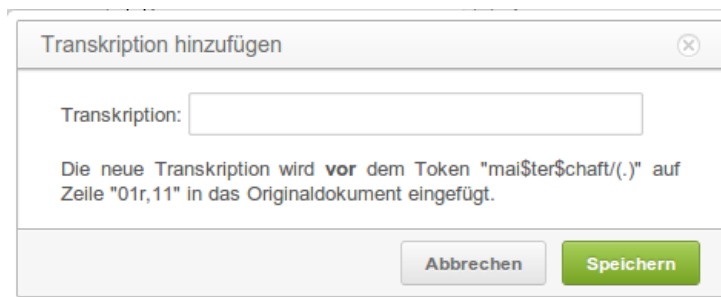


Abbildung 10: Hinzufügen eines Tokens

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 9a zeigt einen Ausschnitt von drei Zeilen aus dem Editor; Abbildung 9b zeigt das Fenster, das sich öffnet, wenn eine beliebige dieser drei Zeilen bearbeitet wird. Die Transkription „mai\$ter\$chaft/(.)“ wird für die Annotation in die drei separaten Token „mai\$ter\$chaft“, „/“ und „(.)“ zerlegt, während die Bearbeitung dieser Transkription jedoch als Ganzes erfolgt.

Im Editorfenster aus Abbildung 9b 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 nicht akzeptiert; die entsprechende Fehlermeldung des Check-Skripts wird dann angezeigt und die Änderung wird nicht vorgenommen.

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 eingegeben werden kann (s. Abbildung 10). Dabei gelten dieselben Regeln wie für das Bearbeiten einer Transkription (siehe oben).

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 Hinweisenfenster 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$tlischen Thiroli$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 sich die neue Transkription nicht am Zeilenende befindet.

Behelfen Sie sich **in keinem Fall**, indem Sie den Zeilenumbruch weglassen! 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“ (mit Zeilenumbruch zwischen den Wörtern).
2. Löschen Sie nun einfach das zweite „schah“.

5 Tastaturbefehle

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.)

Innerhalb von Textfeldern (außer Lemma) können die **Pfeiltasten nach oben/unten** außerdem zum Navigieren zwischen den Zeilen benutzt werden.

Mit **Strg+F** öffnet sich das Suchfenster (s. Abs. 3.3).

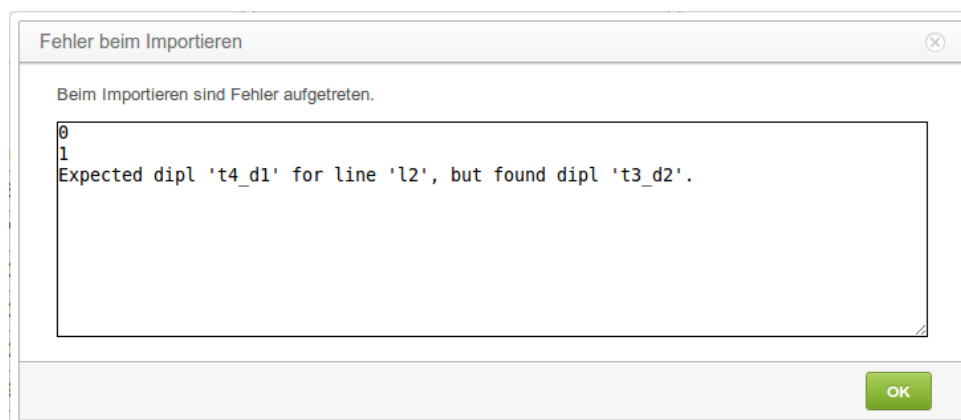


Abbildung 11: Eine Fehlermeldung beim Import

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 11 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).
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!