

Das OCR-Programm

Übersicht

Das hier vorgestellte OCR-Programm ist auf die Digitalisierung umfangreicher Sanskrit-Texte (z.B. ganzer Bücher) ausgerichtet, die in Devanagari gedruckt sind. Es ist im Gegensatz zu handelsüblichen OCR-Programmen für die lateinische Schrift NICHT für die Digitalisierung geringer Textmengen in unbekanntem Fonts gedacht.

Der Digitalisierungsvorgang unterteilt sich in folgende Schritte:

- Import der Bildvorlage eines gedruckten Texts, z.B. über einen Scanner.
- Vorbereitung der Texterkennung: Verbesserung des gedruckten Bildes, Umwandlung in ein Schwarz-Weiss-Bild, Segmentierung der Seite.
- Texterkennung mit der Möglichkeit, unbekannte Zeichen zu trainieren.
- Korrektur und Export des erkannten Textes in ein externes Schreibprogramm.

ACHTUNG: Wenn Sie beabsichtigen, einen umfangreichen Text zu digitalisieren, sollten Sie zunächst anhand der ersten Seite mit den Einstellungen zu Bildauflösung und Kontrast experimentieren, bevor Sie daran gehen, den gesamten Text zu digitalisieren. Details finden Sie in den Kapiteln „Bildvorlagen importieren“ und „Arbeiten mit Stapeldateien“.

Bildvorlagen importieren

Unter dem Begriff „Bildvorlagen“ sind digitale Bilder von gedrucktem Text zu verstehen, dessen Inhalt vom Programm erkannt werden soll. Sie können Bildvorlagen auf zwei Arten in das Programm importieren:

- Wenn Sie Bildvorlagen auf ihrem Rechner gespeichert haben, können sie wie in einem Zeichenprogramm geöffnet werden.
- Gedruckte Vorlagen können über die Scan-Schnittstelle direkt in das Programm importiert werden.

Im folgenden werden die notwendigen Schritte für beide Fälle beschrieben.

Öffnen eines gespeicherten Bildes

- Mit dem Menübefehl „Datei“ -> „Vorlage öffnen ...“ öffnet sich ein typischer Windows-Dialog.
- Wählen Sie im Dialog im Feld „Dateityp“ die Option „Bilder“ aus.
- Wählen Sie das gewünschte Bild und schliessen Sie den Dialog mit „OK“.

Einscannen einer Vorlage

- *Vergewissern Sie sich zunächst, dass ein Scanner an Ihren Computer angeschlossen und dieser Scanner betriebsbereit ist. Der Scanner muss über einen 32-bit-TWAIN-Treiber ansprechbar sein (Standard auf den meisten modernen Rechnern).
- *Falls mehrere Scanner oder z.B. eine Digitalcamera auf Ihrem Rechner benutzt werden, wählen Sie den gewünschten Scanner über den Menübefehl „Datei“ -> „TWAIN-Quelle ...“ aus.
- Wenn der gewünschte Scanner aktiviert ist, betätigen Sie den Befehl „Datei“ -> „Bildvorlage scannen ...“.
- Im jetzt erscheinenden Dialogfeld Ihres Scan-Programms müssen Sie einige **Einstellungen** festlegen:
 - **Farbformat:** Importieren Sie die Vorlage als Graustufen-Bild, d.h. weder die Option „Farbe“ o.ä. noch die Option „SW / Schwarz-Weiss“ o.ä. dürfen aktiviert sein.
 - **Auflösung:** Stellen Sie die Scan-Auflösung auf mindestens 300 dpi ein. Für kleine Schrifttypen ist meist eine Auflösung von 450 dpi empfehlenswert. Experimentieren Sie zunächst mit einer Seite aus der Druckvorlage, bis Sie einen optimalen Wert für die Vorlage gefunden haben (s.a. „Die Bildvorlage vorbehandeln“).
 - **Helligkeit/Kontrast:** Wie bei der Einstellung der Auflösung empfehlen sich hier einige Experimente, bevor Sie mehrere Seiten aus einer Textvorlage einscannen. In Indien gedruckte Bücher sind häufig kontrastarm im Satz; ein Heraufsetzen des Kontrasts vor dem Scannen kann die Erkennungsergebnisse verbessern. Allerdings steht auch innerhalb der Programmoberfläche eine Funktion zur Kontrastverbesserung zur Verfügung, die speziell auf Bilder von gedrucktem Text ausgerichtet ist.
- Scannen Sie jetzt das Bild ein. Es erscheint im Hauptfenster des Programms.

Die Bildvorlage vorbehandeln

Vor dem Beginn der Texterkennung muss das Bild vorbehandelt werden. Die Vorbehandlung umfasst mehrere Schritte (*: Schritt ist optional):

- *Drehen der Vorlage
- *Verbesserung der Kontrasts
- Markieren der Textblöcke
- Segmentierung der Textblöcke

Über das Menü „Ansicht“ können Sie die Vorlage übrigens gemäss Ihren Vorstellung ein- und auszoomen.

*Drehen der Vorlage

Bei Bedarf kann die Bildvorlage mit einem der Drehbefehle (Menü „Rotieren“) um 90 bzw. 180 Grad gedreht werden

*Kontrastverbesserung

Die Kontrastverbesserung (Menü „OCR-Funktionen“ -> „Kontrast verstärken“) reagiert auf Bildbereiche, in denen sich wahrscheinlich Text befindet. Nur in diesen Bereichen wird eine Kontrastverstärkung durchgeführt, woraus sich ein „Flickermuster“ ergibt.

Markieren der Textblöcke

Das Programm besitzt keinen Seitensegmentierungs-Algorithmus, weshalb Sie die Textbereiche manuell markieren müssen.

ACHTUNG: Alle Textbereiche müssen unbedingt **VOR** Beginn der Segmentierung markiert sein, da nur so die korrekte Ausrichtung der Zeilen garantiert ist. Wenn auf einer Seite z.B. drei Textbereiche vorhanden sind, Sie aber zunächst nur den ersten einlesen wollen, müssen Sie trotzdem zuerst alle Bereiche markieren und die überzähligen vor dem Erkennungsprozess löschen (Menübefehl „Cursor“ -> „Textblöcke löschen“; mit dem Cursor auf die unerwünschten Blöcke klicken).

Gehen sie wie folgt vor:

- **Auswahl des Cursors:** Der Cursor zum Markieren der Textblöcke – ein kleines Kreuz - ist standardmässig bei Programmstart aktiviert. Er kann auch über den Menübefehl „Cursor“ -> „Textblöcke festlegen“ geladen werden.
- **Umrahmen Sie die Textbereiche** auf der Seite. Wenn ein Textblock nach dem Aufziehen in seiner Grösse verändert werden soll, muss der Cursor auf dem Rand des Textblocks positioniert werden. Der Cursor verändert seine Gestalt jetzt zu einem Pfeil, mit dem Sie den Textblock in die richtige Position ziehen können.

Achten Sie beim Festlegen der Textbereiche auf die richtige Reihenfolge, die Sie an einer Zahl links oben im aufgezogenen Textblock erkennen. Das Programm liest die Textblöcke in der von Ihnen angegebenen Reihenfolge ein, d.h. eine falsche Numerierung führt zu Konfusion im erkannten Text. Wenn die Textblöcke falsch numeriert sind, können Sie sie mit einem speziellen Cursor (Menü „Cursor“ -> „Textblöcke umnumerieren“) in die richtige Reihenfolge bringen. Klicken Sie dazu mit diesem Cursor in der richtigen Reihenfolge auf die Textblöcke. Die Numerierungen sollten am Ende der korrekten Reihenfolge entsprechen.

Segmentierung der Textblöcke

Nachdem alle Textblöcke markiert sind, wandelt das Programm das Graustufen- in ein Schwarz-Weiss-Bild um und zerlegt die Textbereiche in Zeilen und einzelne Zeichen. Der hierfür verwendete Algorithmus ist noch nicht völlig ausgereift, weshalb er in zwei Variationen angeboten wird („Reinigen“ bzw. „Reinigen mit Standardabweichung“). Nach eigenen Erfahrungen liefert die zweite Version („... mit Standardabweichungen“) in den meisten Fällen bessere Ergebnisse, besonders, wenn bei einer eingescannten Textvorlage grosse dunkle Bereiche in der Nähe der Mittelfalz des Buches auftreten.

Unregelmässigkeiten in der Bildvorlage („Salz und Pfeffer“, d.h. viele kleine isolierte schwarze Punkte) können über eine separate Reinigungsfunktion (Menü „OCR-Funktionen“ -> „Reinigen (schlechte Vorlagen“)) nach dem ersten Segmentierungsprozess entfernt werden.

Nach der Segmentierung liegt die Bildvorlage als SW-Bild vor, in dem die Zeichen gemäss ihrer Position auf drei Arten markiert sind:

- Rote Umrandung: Zeichen in der Hauptzeile;
 - Blaue Umrandung: Zeichen aus der Oberlänge (v.a. Vokalisierungen)
 - Grüne Umrandung: Zeichen aus der Unterlänge
- Prüfen Sie bei den ersten Seiten einer neuen Vorlage unbedingt die Segmentierungsqualität. Eine Auswahl typischer Fehler in diesem Stadium: Zeichen der Oberlänge/Unterlänge werden nicht abgetrennt. So wird z.B. übergeschriebenes „i“ mit dem zugehörigen Laut aus der Hauptzeile verschmolzen
 - Mehrere Zeichen der Hauptzeile sind zu einem grossen Zeichen verschmolzen.
 - Zeichen der Hauptzeile werden zu stark segmentiert, d.h. fälschlich in mehrere Zeichen aufgeteilt. **ACHTUNG:** Die Nagari-Zeichen „ga“, das retroflexe n und das palatale s werden pathologisch falsch aufgeteilt. Das Programm besitzt für diese Spezialfälle einen internen Korrekturmechanismus, so dass diese falschen Segmentierungen nicht als Fehler zu werten sind.

ACHTUNG: Die Qualität der Segmentierung beeinflusst in erheblichem Mass den Erkennungsprozess. Wenn in diesem Stadium zu viele Fehler auftreten, sollten Sie das Bild mit veränderten Einstellungen (z.B. Erhöhung der Auflösung, Verstärkung des Kontrasts, ...) nochmals importieren. Experimentieren Sie mit der ersten Seite eines neuen Textes solange, bis vertretbare Resultate bei der Segmentierung erzielt werden, und merken Sie sich die Scan-Einstellungen (Auflösung, Kontrast, Helligkeit) für diese Seite. Erst jetzt sollten Sie umfangreichere Mengen aus der Druckvorlage einscannen.

Text erkennen

Das Programm verfügt über einen vortrainierten Zeichensatz, der die wichtigsten Nagari-Zeichen mit relativ grosser Genauigkeit erkennt („Omnifont“). Trotzdem wird in den meisten Fällen ein Training auf eine neue Druckvorlage erforderlich sein. Im folgenden wird zunächst die einfachste Option – Erkennung nur mit Omnifont – und danach der Trainingsprozess beschrieben.

Erkennung nur mit Omnifont

- Aktivieren Sie die Omnifont-Option (Menü „OCR-Funktionen“ -> „Omnifont benutzen“).
- Betätigen Sie den Befehl „OCR-Funktionen“ -> „Texterkennung“.

- Der Text wird erkannt und im unteren Bereich des Hauptfensters ausgegeben.

Erkennung mit Training

Wenn Sie mit den Ergebnissen der ersten Prozedur nicht zufrieden sind, müssen Sie einen speziellen Zeichensatz erstellen und trainieren.

**1. Schritt: Festlegen des Standardordners*

(ACHTUNG: Dieser Schritt muss nur beim ersten Trainingsvorgang durchgeführt werden)

- Erstellen Sie mit einem externen Dateiverwaltungsprogramm (z.B. WindowsExplorer) einen Ordner (z.B. „c:\Zeichensätze“), und merken Sie sich, wo sich der Ordner auf Ihrer Festplatte befindet.
- Wählen Sie den Menübefehl „Zeichensatz“ -> „Standardordner für Zeichensätze“. Markieren Sie im jetzt erscheinenden Dialog den eben erstellten Ordner und schliessen Sie den Dialog.

Neue Zeichensätze werden von jetzt an in diesem Standardordner gespeichert. Über die Liste „trainierbarer Zeichensatz“ im Werkzeugfenster können Sie schnell auf den Inhalt dieses Ordners zugreifen.

2. Schritt: Erstellen des trainierbaren Zeichensatzes

- Wählen Sie den Menübefehl „Zeichensatz“ -> „Neuer Zeichensatz“. Ein Dialog erscheint.
- Geben Sie dem Zeichensatz einen schönen Namen (Feld „Name des Zeichensatzes“; Vorschlag: Name des Textes / der Reihe, in der der Text erschienen ist) und schliessen Sie den Dialog.

ACHTUNG: Erstellen Sie für jeden neuen Schriftsatz einen neuen Zeichensatz. Das Mischen verschiedener Schriftsätze in einem Zeichensatz kann die Erkennungsqualität erheblich verringern!

3. Schritt: Training aktivieren

Aktivieren Sie die Option „OCR-Funktionen“ -> „Training an / aus“. Es wird empfohlen, neben der Trainingsoption auch den Omnifont zu aktivieren.

4. Schritt: Texterkennung

Betätigen Sie den Befehl „OCR-Funktionen“ -> „Texterkennung“. Die Texterkennung beginnt.

Sobald das Programm ein Zeichen nicht sicher identifizieren kann, wird während des Erkennungsprozesses der **Trainings-Dialog** eingeblendet. Dieses Fenster verfügt über zwei wichtige Bereiche:

- Links ist ein Ausschnitt aus der Bildvorlage eingeblendet, in dem das unidentifizierte Zeichen rot umrahmt ist. Die Position des Rahmens kann in der aktuellen Version nicht verändert werden.
- Rechts oben befindet sich ein Textfeld, in dem in den meisten Fällen eine Transkription vorgeschlagen ist. Korrigieren Sie nötigenfalls die angegebene Transkription, und betätigen Sie dann die Schaltfläche „Zeichen trainieren“. Das Zeichen wird jetzt mit der angegebenen Transkription an den trainierbaren Zeichensatz angehängt, sofern eine Maximalmenge (momentan 30) von Zeichen dieser Klasse im Zeichensatz nicht überschritten ist (d.h.: Wenn das Zeichen „ya“ schon in 30 Varianten vorhanden ist, wird die 31. zwar in den erkannten Text eingesetzt, aber nicht mehr gespeichert).

Weitere Befehle im Trainings-Dialog:

- „Zeichen speichern“: Dieser Befehl bewirkt grundsätzlich dasselbe wie der Befehl „Zeichen trainieren“. Allerdings wird das Zeichen ohne Rücksicht auf die Maximalmenge gespeichert.
- „Ignorieren“: Mit diesem Befehl überspringen Sie das markierte Zeichen. Es wird damit nicht in den Ausgabertext aufgenommen. Dieser Befehl kann z.B. bei fehlerhaften Segmentierungen oder bei irrelevanten Zeichen (Klammern, Sternchen, ...) eingesetzt werden.
- „Training beenden“: Hiermit wird das Training beendet. Die Texterkennung wird für den Rest der Seite ohne Training durchgeführt.
- „Texterkennung beenden“: Die Texterkennung wird an der aktuellen Position beendet.
- „Zeichensatz bearbeiten“: Dieser Befehl öffnet den Zeichensatz-Dialog, in dem Sie z.B. Fehleingaben während der Trainingsphase korrigieren können.

Erkannten Text bearbeiten

Nach dem Ende der Texterkennung wird der erkannte Text in den unteren Bereich des Hauptfensters geladen, wo Sie ihn bearbeiten und korrigieren können.

Das Programm tritt dazu in den sogenannten „Korrekturmodus“ ein. Dabei wird die Position des Cursors im erkannten Text mit dem Bereich der Bildvorlage koordiniert, aus dem der Text stammt. Wenn Sie den Cursor im Textfeld bewegen, wird die Bildvorlage automatisch auf die entsprechende Position bewegt, sodass Sie den erkannten Text mit dem gedruckten vergleichen können. Sie können diesen Modus über den Menübefehl „OCR-Funktionen“ -> „Korrekturmodus“ an- und ausschalten.

Der erkannte und korrigierte Text kann in die Zwischenablage (Menü „Erkannter Text“ -> „in die Zwischenablage“) exportiert und von dort in jedes handelsübliche Schreibprogramm eingefügt werden. Achten Sie darauf, dass in dem Schreibprogramm der richtige Font aktiviert ist.

Arbeiten mit Stapeldateien

Wie in der Übersicht bemerkt, ist das Programm auf das Einlesen umfangreicher Texte ausgerichtet. Wenn Sie z.B. ein Buch von mehreren hundert Seiten digitalisieren wollen, wäre es unpraktisch, die Einstellungen zur Texterkennung bei jeder Seite von neuem eingeben zu müssen. In diesem Fall sollten Sie eine Stapeldatei anlegen.

In einer Stapeldatei wird eine Reihe von Bildvorlagen zusammen mit den globalen Einstellungen (z.B. zum trainierbaren Zeichensatz) gespeichert. Der Inhalt der Stapeldatei wird im Werkzeugfenster angezeigt. Wenn Sie einen Eintrag in der oberen Liste aktivieren, wird die entsprechende Vorlage in das Hauptfenster geladen und kann dort bearbeitet werden.

Der Inhalt einer einmal angelegten Stapeldatei wird automatisch bei Programmschluss gespeichert.

Erstellen einer neuen Stapeldatei

- Bei Programmstart wird als Standard eine neue leere Stapeldatei angelegt, mit der Sie sofort arbeiten können (d.h. der Befehl „Datei“ -> „Neue Stapeldatei“ muss nicht aufgerufen werden).
- Wenn Sie mit einer Stapeldatei arbeiten und eine neue leere Stapeldatei erzeugen wollen, wählen Sie dazu den Menübefehl „Datei“ -> „Neue Stapeldatei“.

Öffnen einer Stapeldatei

Wählen Sie den Befehl „Datei“ -> „Vorlage öffnen ...“ und geben Sie im folgenden Dialog den Pfad zu der Stapeldatei an. Sie können die Bildvorlage in dieser Stapeldatei jetzt bearbeiten oder ihr z.B. neue Vorlagen über den Scanner hinzufügen.

Löschen von Vorlagen aus der aktuellen Stapeldatei

Wenn Sie eine Textseite digitalisiert haben, werden Sie sie vermutlich aus der Stapeldatei und von Ihrer Festplatte entfernen wollen.

Wählen Sie dazu den Menübefehl „Datei“ -> „Akt. Vorlage löschen“. Nach einer Rückfrage wird die Bildvorlage aus der Stapeldatei und von Ihrem Rechner entfernt.

Standardordner zum Speichern der Bilder

Für jede Bildvorlage in Ihrer Stapeldatei wird eine eigene temporäre Datei angelegt, die auf Ihrer Festplatte gespeichert wird. Mit dem Befehl „Datei“ -> „Standardordner zum Speichern der Bilder ...“ können Sie den Ordner festlegen, in dem diese Dateien gespeichert werden sollen. Der Befehl kann z.B. bei partitionierten Festplatten

eingesetzt werden, damit das Programm die temporären Dateien in freien Partitionen ablegen kann.

Arbeiten mit Zeichensätzen

Wenn Sie mit einem trainierbaren Zeichensatz arbeiten, kann die Erkennungsqualität in erheblichem Umfang von der Güte Ihrer Eingabe abhängen. Wenn Sie z.B. – was nicht geschehen möge – zahlreiche Zeichen während des Trainingsprozesses falsch klassifizieren, werden diese Fehler in künftigen Erkennungsvorgängen reproduziert. Daher empfiehlt es sich, die Zeichen des trainierbaren Zeichensatzes gelegentlich zu überprüfen:

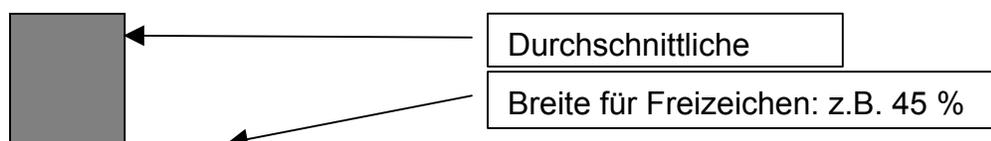
- Wählen Sie den Menübefehl „Zeichensatz“ -> „Zeichensatz bearbeiten“. Der Zeichensatz-Dialog, der schon beim Erstellen eines neuen Zeichensatzes erschien, wird geöffnet.
- Auf der rechten Seite des Fensters werden Informationen über die Zeichenklassen des aktuellen Zeichensatzes angezeigt: Das Feld „Transkription der Zeichen“ gibt die Transkription für alle dargestellten Zeichen an, das Feld „bei Erkennung ignorieren“, ob die Zeichenklasse in den Erkennungsprozess einbezogen werden soll.
- Zwischen den gespeicherten Zeichenklassen können Sie mit den Pfeiltasten („<<“ usw.) oder über die Suchliste navigieren. Wenn die Suchliste blau markiert ist, können Sie auch mit den Pfeiltasten nach oben/unten zwischen den Klassen wechseln.

Korrektur von Zeichenklassen

Wenn Sie z.B. bemerken, dass in der „ma“-Klasse aus Versehen ein „bha“ eingeschlossen ist, markieren Sie dieses Zeichen mit dem Cursor (es ist jetzt rot gerahmt). Betätigen Sie dann die Taste „Markiertes Zeichen löschen“ oder die ENTF-Taste. Das Zeichen wird aus der aktuellen Klasse entfernt.

ACHTUNG: Sie könnten auf die Idee kommen, das falsche Zeichen zu markieren und dann die Transkription im Feld „Transkription der Zeichen“ zu verändern. In diesem Fall wird sämtlichen sichtbaren Zeichen (d.h. der gesamten Zeichenklasse) die neue Transkription zugewiesen!!

Neben der Korrektur von Zeichenklassen kann auch der Segmentierungsprozess über dieses Fenster beeinflusst werden. Das Feld „Seitenverhältnis bei Leerstellen“ gibt an, wieviel Prozent einer durchschnittlichen Zeichenhöhe auf ein Freizeichen in einer Zeile hindeuten (vgl. das folgende Bild).



Bei eng gedrucktem Text sollte dieser Wert eher niedrig (z.B. 30%), bei weit gedrucktem eher hoch (>> 50%) gewählt werden.

Wenn Sie mit Ihren Eingaben zufrieden sind, schliessen Sie den Dialog mit „Übernehmen“.

Zeichensätze wiederverwerten

Angenommen, Sie arbeiten schon längere Zeit mit dem Programm und haben Text aus Büchern mit fünf verschiedenen Zeichensätzen digitalisiert, so dass jetzt fünf verschiedene trainierbare Zeichensätze vorhanden sind. Sie sind dabei, ein neues Buch zu digitalisieren, das in einem Schriftsatz gedruckt ist, den Sie schon trainiert haben (Beispiel: Sie haben schon ein Buch aus der Reihe „Kashmir Series of Texts and Studies“ eingelesen und wollen jetzt einen weiteren Band dieser Reihe digitalisieren). In diesem Fall müssen Sie keinen neuen Zeichensatz anlegen und von vorne trainieren, sondern können auf den schon bestehenden zurückgreifen. Gehen Sie dazu wie folgt vor:

- Wenn sich der gesuchte Zeichensatz im Standardordner für Zeichensätze befindet, können Sie ihn aus der Liste „trainierbarer Zeichensatz“ im Werkzeugfenster auswählen.
- Wenn sich der Zeichensatz in einem anderen Verzeichnis befindet, wählen Sie den Menübefehl „Zeichensatz“ -> „Zeichensatz öffnen ...“ und geben Sie im folgenden Dialog den Pfad des Zeichensatzes an.

Der gewählte Zeichensatz ist jetzt einsatzbereit und kann weiter trainiert werden.

Zeichensätze löschen

Grundsätzlich sollten Sie einen korrekten Zeichensatz nicht löschen: Sie könnten ihn bei weiteren Digitalisierungen nochmals verwenden.

Wenn Sie einen Zeichensatz unbedingt löschen wollen, wählen Sie den Befehl „Zeichensatz“ -> „Zeichensatz löschen“. Der aktuell geladene Zeichensatz wird damit unwiederbringlich von Ihrer Festplatte entfernt.

Das Sprachanalyse-Programm

Mit Texten arbeiten

Eine Textpassage laden

Wenn Sie eine Passage aus einem Text laden wollen, der in der Datenbank gespeichert ist, gehen Sie wie folgt vor:

- Aktivieren Sie die Registerkarte „Texte“ auf der rechten Seite des Programmfensters.
- Wählen Sie mit Doppelklick den gesuchten Text aus der Liste „Texte“ unten in der Registerkarte aus. In das Fenster „Textstruktur“ wird jetzt die Gliederung des Textes in Form eines Baumdiagramms geladen. In diesem Diagramm werden folgende Symbole benutzt:
- Ein Punkt neben dem Namen eines Kapitels zeigt an, dass es sich hier um eine normale Textebene handelt.
- Ein „C“ neben dem Namen des Kapitels bedeutet, dass diese Textebene eine Kommentarebene zu einem Vers / einer Strophe aus dem übergeordneten Kapitel ist. Die Nummer der kommentierten Strophe wird neben dem Namen des Kommentars angezeigt.
- Markieren Sie das gesuchte Kapitel. Der Name des Kapitels und das entsprechende Symbol erscheinen jetzt rot. Gleichzeitig werden in den Textfeldern „Strophe“ unter dem Baumdiagramm die erste und letzte Strophe des gewählten Kapitels angezeigt. Wenn das Kapitel z.B. 100 Strophen besitzt (-> die Werte „1“ und „100“ werden angezeigt), Sie sich aber nur für die Strophen 10 bis 25 interessieren, geben Sie diese Werte in die Felder ein.
- Betätigen Sie die Schaltfläche „Text anzeigen“ bzw. die ENTER-Taste. Der Text wird in das Hauptfenster des Programms geladen.

Die grammatikalische und lexikalische Auflösung von Textpassagen

Sobald ein Text in das Hauptfenster geladen ist, sind die grammatikalischen und lexikalischen Informationen zu der geladenen Passage (falls vorhanden) verfügbar.

Wenn Sie mit dem Mauszeiger in ein Wort des geladenen Textes klicken, verändert sich der Inhalt zweier Listen im unteren Bereich des Hauptfensters:

- In der grossen Auflösungsliste (links) wird eine Zeile gefüllt, falls das ausgewählte Wort im Text analysiert worden ist. In dieser ersten Zeile finden Sie die Auflösung für das gewählte Wort. Die Lemmata, aus denen das Wort besteht, sind durch Bindestriche voneinander getrennt. Standardmässig ist das erste Wort aktiviert (= rot). Für dieses Wort werden in die zweite Liste (rechts im unteren Bereich) die Bedeutungen aus dem programminternen Lexikon geladen.

- Sobald Sie mit dem Cursor das nächste Wort in der Auflösungsliste markieren, werden die Bedeutungen für dieses Wort geladen usw. Sie können sich innerhalb einer Auflösung auch mit den Pfeiltasten nach rechts und links bewegen. Wenn dabei das Ende einer Auflösung erreicht wird, springt das Programm automatisch zum Beginn der vorhergehenden/nächsten Wortes, für das die beschriebene Prozedur wiederholt wird.

Lexikon und Wortsuche

Das programminterne Lexikon erlaubt im Gegensatz zu bisher üblichen stringbasierten Suchvorgängen die Suche nach Grundformen von Wörtern unabhängig von ihren im Text realisierten morphologischen Varianten.

Dabei kann das Lexikon nach Sanskrit-Wörtern (Eingabefeld „Lemma“) oder ihren Bedeutungen (Einfagefeld „Bedeutung“) durchsucht werden. Beide Suchvorgänge können mit Wildcards durchgeführt werden.

Die Ergebnisse der Lexikon-Suche werden in der Liste links im Lexikon-Fenster angezeigt. Wenn Sie einen Eintrag in dieser Liste markieren, werden in der Liste „Bedeutungen“ die Bedeutungen des Wortes angezeigt.

Suche nach Sanskrit-Wörtern:

Geben Sie im Textfeld „Lemma“ den relevanten Stamm des gesuchten Wortes ein und drücken Sie die ENTER-Taste.

Suche nach Bedeutungen:

Geben Sie im Textfeld „Bedeutung“ die gesuchte Bedeutung ein, und drücken Sie die ENTER-Taste. Es werden alle Wörter ausgegeben, denen die gesuchte Bedeutung zugeordnet ist.

Wildcards für die erweiterte Suche

Mit Wildcards – Platzhalterzeichen für flexible Zeichenausdrücke - können Sie die Flexibilität eines Suchvorgangs erhöhen. Eine Suche mit Wildcards lässt sich z.B. bei der Edition einer Handschrift einsetzen: Wenn innerhalb einer Zeichenfolge einzelne aksharas nicht erkennbar sind, kann an diesen Positionen der Platzhalter * eingesetzt werden.

Wildcard Funktion

?

Platzhalter für ein beliebiges einzelnes Zeichen. – **ACHTUNG:**
Diphthonge und Aspirata werden als EIN Zeichen aufgefasst.

Beispiel: r?ma kann stehen für rama, roma, usw.

* Platzhalter für kein oder mehrere Zeichen.

Beispiele:

T*r kann stehen für Tür, Tor, Teer, Tr, usw.

bud?acar* (dh wird nach Nagari-Konvention als ein Laut aufgefasst - > ?) liefert u.a. buddhacarita, buddhacaritrya, buddhacarya

□ Die eckigen Klammern zeigen an, daß die in ihnen eingeschlossenen Zeichen alternativ an dieser Position auftreten können.

Beispiele:

T[üo]r liefert Tür und Tor, aber nicht Teer usw.

r[äo]m liefert u.a. rāma, roma, roman, rāmī

ACHTUNG: Wenn Sie bei der Suche den Laut , (Apostroph; Shift+#) einsetzen, meldet das Programm einen Fehler. Vermeiden Sie deshalb diesen Laut.

Fundstellen suchen

Die Fundstellen eines Wortes aus dem Lexikon zeigen Sie folgendermassen an:

Markieren Sie ein Wort in der Lexikon-Liste.

Wählen Sie in der Liste „Umfang der Suche“ den Suchumfang aus. **ACHTUNG:** Wenn Sie die Optionen „aktueller Text“ oder „aktuelles Kapitel“ auswählen, aber momentan kein Text aktiviert ist, erzeugt das Programm eine Fehlermeldung und stellt den Standard-Suchumfang ein. Zum Aktivieren eines Textes/Kapitels lesen Sie den Abschnitt „Eine Textpassage laden“!

Betätigen Sie die Schaltfläche „Suchen“. Die Fundstellen – falls vorhanden – werden in der Liste „Fundstellen“ angezeigt.

Wenn Sie einen Eintrag in der Fundstellen-Liste mit dem Cursor markieren und die linke Maustaste gedrückt halten, erscheint nach wenigen Augenblicken ein Popupfenster mit einer Vorschau auf den Text. Um den Text in das Hauptfenster zu laden, sollten Sie doppelt auf den Eintrag in der Liste klicken.

Lexikoneinträge erstellen und bearbeiten

Wenn sich im Verlauf der Textanalyse ein Wort nicht (-> Wort neu erstellen) oder nicht mit den gewünschten grammatikalischen Klassen (-> Wort bearbeiten) im Lexikon findet, die Textanalyse einen String also nicht auflösen kann, müssen Sie die Lexikon-Datenbank um das gesuchte Wort erweitern.

Einen Lexikoneintrag neu erstellen

- Betätigen Sie in der Registerkarte „Lexikon“ die Schaltfläche „Neuer Lexikoneintrag“. Der Dialog „Lemma erstellen/bearbeiten“ wird geöffnet.

- Tragen Sie im Feld „relevanter Stamm“ den relevanten Stamm des Wortes ein. Während des Schreibens werden in die Liste „ähnliche Wörter“ alle im Lexikon vorhandenen Einträge geladen, die dem Wort ähneln (eine Vorsichtsmassnahme gegen doppelte Einträge).

Nominalformen

- Wenn es sich bei dem Wort um ein Nomen oder ein Indeklinabile handelt, aktivieren Sie die Registerkarte „Nomina und Indeklinabilia“.
- Wählen Sie aus der Liste „vorhandene Kategorien“ die Nominalklasse(n) aus, die dem Wort zugeordnet werden soll(en). Einem Wort können mehrere Nominalklassen zugeordnet sein. Achten Sie dabei auf die lexikalische Kohärenz Ihrer Einträge!
- Betätigen Sie jetzt die Schaltfläche „Kat. zuordnen“. Die ausgewählte Nominalklasse wird dem Wort zugeordnet und erscheint in der Liste „zugeordnete Kategorien“.

Verbalformen

Die Erstellung von Verbalformen erfordert eine grössere Anzahl von Schritten, damit die neuen Angaben mit der auf die indischen Grammatiker gestützten Datenbankstruktur kohärent ausfallen. Grundsätzlich ist zwischen Basisverben (= Verben, die einer der ersten neuen traditionellen Verbalklassen entstammen und kein Präfix besitzen; dieser Fall sollte nur noch sehr selten auftreten), Verben der Sekundärkonjugation (= Verben, die nicht den ersten neuen Präsensklassen entstammen, aber kein Präfix besitzen) und präfigierten Verben zu unterscheiden.

Tragen Sie in jedem Fall zuerst die Grundform des Verbs in das Feld „relevanter Stamm“ ein. Aktivieren Sie dann die Registerkarte „Verben“.

Basisverben

- Wählen Sie aus der Liste „Bindevokal“ den passenden Typ.
- Wählen Sie aus der Liste „Präsensklassen“ die richtige Präsensklasse, und betätigen Sie die Schaltfläche „Pr.klasse zuordnen“. Die neue Präsensklasse erscheint jetzt in der Liste „Zugeordnete Präsensklassen“.
- Wiederholen Sie – wenn nötig – den letzten Schritt, bis dem Verb alle nötigen Präsensklassen zugeordnet sind.

Verben der Sekundärkonjugation

- Führen Sie zunächst die Schritte aus, die bei der Erstellung eines Basisverbs nötig sind.
- Betätigen Sie jetzt die Schaltfläche „Sekundärkonjugation“. Im folgenden Dialog müssen Sie das Basisverb markieren, von dem das sekundäre abgeleitet ist.

Beispiel: Sie wollen den Kausativ *gāmay* als Sekundärbildung vom Basisverb *gam* erstellen.

- Tragen Sie zuerst die Form *gāmay* im Feld „relevanter Stamm“ ein.
- Aktivieren Sie die Registerkarte „Verben“.
- Wählen Sie als Bindevokaltyp „seṭ“.
- Wählen Sie aus der Liste „Präsensklassen“ den Eintrag „10. Klasse (= Kausativ)“.
- Betätigen Sie die Schaltfläche „Pr.klasse zuordnen“.
- Betätigen Sie die Schaltfläche „Sekundärkonjugation ...“.

- Wählen Sie im folgenden Dialog das Basisverb *gam* aus und schliessen Sie den Dialog mit „Übernehmen“.

Präfigierte Verben

Bei der Erstellung von präfigierten Verben müssen Sie dem Programm die Lautgestalt des Präfixes und das entsprechende Basisverb mitteilen. Am Beispiel des Wortes *anugam*:

- Tragen Sie zunächst die Form *anugam* in das Textfeld „relevanter Stamm“ ein.
- Aktivieren Sie die Registerkarte „Verben“.
- Tragen Sie auf der rechten Seite der Registerkarte im Feld „Präfixe“ das Präfix *anu* ein.
- Wählen Sie aus der Liste „Bilden nach ...“ das Basisverb *gam* aus.
- Betätigen Sie die Schaltfläche „Zuordnen ...“. Im jetzt erscheinenden Dialog sind alle Präsensklassen verzeichnet, die dem Basisverb zugeordnet sind. Wählen Sie mindestens eine der Präsensklassen aus, und schliessen Sie den Dialog mit „Übernehmen“.

Einen Lexikoneintrag bearbeiten oder löschen

- Suchen Sie den gewünschten Eintrag im Lexikon. Markieren Sie ihn in der Liste der gefundenen Wörter.
- Betätigen Sie jetzt die Schaltfläche „Bearbeiten/Löschen“. Der Dialog „Lemma erstellen/bearbeiten“ wird geöffnet.
- Veränderung der Lautgestalt des Lexikoneintrags: Bearbeiten Sie das Wort im Feld „relevanter Stamm“, und schliessen Sie den Lexikon-Dialog mit „OK“.
- Einem Wort neue grammatikalische Klassen zuweisen: Wenn das Wort *vajra* – mit identischer Bedeutung – als Maskulinum und Neutrum auftritt, das Lexikon bisher aber nur die Kategorie „a mask.“ verzeichnet, sollten Sie aus der Liste „Vorhandene Kategorien“ die gewünschte weitere Kategorie wählen und dem Wort mit der Taste „Kat. zuordnen“ zuweisen.
- Löschen von grammatikalischen Kategorien: Markieren Sie in der Liste „zugeordnete Kategorie“ die Kategorie, die entfernt werden soll, und betätigen Sie die Taste „Kat. löschen“.
- Löschen eines Lemmas aus dem Lexikon: Wiederholen Sie den vorherigen Schritt so oft, bis alle dem Wort zugeordneten grammatikalischen Kategorien entfernt sind. Das Wort ist jetzt aus dem Lexikon beseitigt.

Bedeutungen erstellen und bearbeiten

Die Bedeutungen jedes Wortes im Lexikon können vom Benutzer frei bearbeitet werden. Dazu steht in der Sprachanalyse-Oberfläche, neben dem Lexikon und bei der Erstellung

neuer Lexikoneinträge ein gesondertes Steuerelement zur Verfügung, das aus einer Liste und mehreren Schaltflächen besteht.

Eine neue Bedeutung erstellen:

- Geben Sie die neue Bedeutung im Textfeld unter der Liste der vorhandenen Bedeutungen ein, und betätigen Sie die Anfügen-Taste (“+”).
- ODER: Betätigen Sie die Taste “Neu”. Ein Dialog erscheint, in dem die gewünschte Bedeutung eingetragen werden kann.

Eine Bedeutung bearbeiten:

Klicken Sie doppelt auf den entsprechenden Eintrag in der Liste. Im jetzt erscheinenden Dialog können Sie die Bedeutung verändern.

Eine Bedeutung löschen:

- Markieren Sie die Bedeutung in der Liste.
- Betätigen Sie die Schaltfläche “Löschen” (“X”). Nach eine Nachfrage wird die Bedeutung aus der Programmdatenbank gelöscht.

Texte grammatikalisch und lexikalisch analysieren („Tagging“)

Auflösungen generieren

Wenn sich im aktuellen Textabschnitt unaufgelöste Lautfolgen (Schrift ist rotbraun <-> schwarz: aufgelöste Lautfolgen) befinden, können Sie mit dem Befehl „Nicht erkannte Wörter analysieren“ die Grammatikanalyse aufrufen. Neu erkannte Wörter werden im Text rot markiert, in der Auflösungsliste unten im Hauptfenster ist in den meisten Fällen eine Reihe von verschiedenen Möglichkeiten angegeben, die Sie jetzt überprüfen sollten.

Auflösungen auswählen

Die Darstellung der Ergebnisse der Sprachanalyse orientiert sich am Aufbau von Sanskrit-Komposita. In der grossen Liste unten im Hauptfenster sind übereinander verschiedene Auflösungsmöglichkeiten für die aktuell markierte Lautfolge aufgelistet. An jeder Position innerhalb eines dieser Komposita treten in den meisten Fällen wiederum eine Reihe von Auflösungsvarianten auf. Am Beispiel der Lautfolge *kāryatvam*:

kārya	tva
<i>kārya (a neutr.)</i>	<i>tva (Marker für Abstraktnomina)</i>
<i>kṛ (Gerundivum)</i>	<i>tvad ("du"; Stammform)</i>
...	...

ka	ārya	tva (s.o.)
	<i>ārya (a mask.)</i>	
	<i>ārya (Adj.)</i>	
	...	

Wenn eine Lautfolge den Bestandteil – kārya - enthält, kann das Programm diesen Ausschnitt – gemäss der oben dargestellten Tabelle - auf verschiedene Arten analysieren. Eine erste Möglichkeit wäre die Form kārya, dann ist z.B. aber auch die Auflösung in kā und ārya möglich. Das Programm betrachtet mangels semantischer Information diese Möglichkeiten zunächst als gleichwertig, erstellt also zwei Listeneinträge (die fett gedruckten Zeilen): Einen mit – kārya (<gramm. Angaben>) - , einen zweiten mit – kā (<gramm. Angaben>) – ārya (<gramm. Angaben>) - . Analog wird für mit dem zweiten bzw. dritten Bestandteil des Kompositums (-tva-) verfahren.

Erschwerend wirkt sich beim ersten Eintrag die Herleitung des Wortes kārya aus, das u.a. als Gerundivum von kṛ abgeleitet werden kann, aber auch als selbstständiger Eintrag (als Neutrum „Objekt“ usw.; und als Adjektiv) im Lexikon vermerkt ist. Und diese Vieldeutigkeit ist leider nicht nur auf diesen einen Eintrag beschränkt! Wenn Sie in einem neu analysierten Text z.B. auf den Listeneintrag kārya doppelt klicken, erscheint eine zweite Liste, in der mit absteigender Wahrscheinlichkeit alle weiteren Herleitungen aufgeführt sind (die kleiner + kursiv gedruckten Zeilen in der Tabelle). Je länger Sie mit einem Text arbeiten, desto treffsicherer wird diese Vorauswahl des Programms.

Wählen Sie aus dieser zweiten Liste die richtige Herleitung aus – sofern sie sich nicht schon an der Spitze der Liste befindet - und schliessen Sie die zweite Liste (Doppelklick auf die richtige Herleitung / Herleitung markieren -> ENTER bzw. #). Die richtige Herleitung steht jetzt oben in der Liste. Stellen Sie anschliessend sicher, daß die aktuellen Herleitungen auch ganz oben in der Auflösungsliste stehen. Wenn nicht, markieren Sie den Eintrag in der Auflösungsliste und betätigen Sie den Knopf „Komp[ositem] aktivieren“ (rechts im Fenster); bzw. klicken Sie mit gedrückter CTRL-Taste doppelt auf den Eintrag. Wiederholen Sie diese Schritte, falls nötig, für alle Einträge und alle Herleitungen.

Auflösungen speichern

Wenn jetzt für alle Lautfolgen des Textabschnitts die korrekten Auflösungen an der Spitze der Auflösungsliste stehen, können Sie den Text speichern (Strg+S, bzw. Befehl aus dem Menü). Deaktivierte Auflösungen und Herleitungen, d.h. alles, was nicht ganz oben in der Liste stand, werden vom Programm vollständig gelöscht.

Navigieren in den Auflösungen; weitere Befehle zur Sprachanalyse

Unter der Voraussetzung, dass im Textfenster eine Auflösung markiert ist, stehen folgende Tastaturbefehle im Auflösungsfenster bereit:

Tastenkombination	Knopf / Alternative	Funktion
Pfeiltasten nach links / rechts		Zur vorigen/nächsten Einzelform in der Auflösung. Die Einzelform wird aktiviert und rot markiert. Befindet sich in der aktuellen Auflösung keine Einzelform mehr, springt das Programm zum nächsten Wort. Rechts werden die Bedeutungen des aktivierten Lemmas angezeigt.
# (wenn eine Einzelform markiert (= rot) ist)	Doppelklick auf Wort	Die Liste mit alternativen Herleitungen wird angezeigt (vgl. oben). Diese Funktion ist nur verfügbar, wenn ein Wort neu analysiert und noch nicht gespeichert wurde.
CTRL+Seite hoch (page up)	„Komp. aktivieren“	Die ausgewählte Auflösung wird in der Liste ganz nach oben gestellt und somit aktiviert; identisch mit Knopf „Kompositum aktivieren“.
DEL (mit Focus auf Auflösungsliste)	„Alle XX (DEL)“	Alle Auflösungen des aktuellen Wortes werden nach Rückfrage gelöscht. Davon betroffen sind auch Einträge in der Datenbank und Spezialbedeutungen, aber nicht die Einträge in ihrer Textgestalt oder die verknüpften Lemmata im Lexikon.
ENTER	„Wort bearbeiten“	Das aktuell markierte (= rote) Wort in der Auflösung bearbeiten. Dazu wird der Dialog „Eintrag erstellen / bearbeiten“ geöffnet, in dem Sie die grammatikalischen Angaben zum ausgewählten Wort verändern können.
TAB (Tabulator)		Wenn der Focus auf dem Auflösungsfenster war, springt er ins Textfeld, wo Sie sofort per Tastatur weiterarbeiten können. Lag er im Textfeld, wird er durch TAB in das Auflösungsfenster gesetzt.
CTRL+ENTER	„Lex. bearbeiten“ „Neu: Kompositum“	Bearbeitet das Lemma im Lexikon, das dem markierten (= roten) Wort zugrundeliegt. Dazu wird der Dialog „Lemma erstellen / bearbeiten“ geöffnet. Öffnet den Dialog „Neue Auflösung erstellen“, mit dessen Hilfe Sie manuell eine völlig neue Auflösung der aktuellen Lautfolge erstellen können (wenn z.B. die automatische Grammatikanalyse nicht die gewünschten Resultate gezeigt hat).

Ein Teil der beschriebenen Befehle ist auch über die Schaltfläche „Aktionen“ zugänglich.

Texte bearbeiten (Textgestalt verändern)

Obwohl sich das Fenster auf den ersten Blick nicht von normalen ®Windows©-Textfeldern unterscheidet, werden Sie bei der Arbeit mit dem Textfeld schnell eine Reihe von Unterschieden bemerken. Die andersartige Funktionalität hängt damit zusammen, daß das Textfeld direkten Zugriff auf die Textdatenbank bietet. Eine durch Freizeichen abgeschlossene Lautfolge, ein Wort, wird in der Datenbank nicht im Fliesstext gespeichert, sondern ist durch einen eigenen Eintrag repräsentiert, mit dem die Auflösung in Einzellemmata, eventuelle Übersetzungen, Spezialbedeutungen, Lesarten und Kommentare verknüpft sind. Beim Laden einer Textpassage fragt das Programm alle zugehörigen Wörter ab, lädt die jeweils verknüpften Informationen und verkettet die Wörter zu dem Gesamttext, den Sie im Textfeld sehen. Deshalb dauert das Laden einer Passage deutlich länger als z.B. in einem Textverarbeitungsprogramm.

Sobald der Text geladen ist, assoziiert das Programm mit jeder Lautfolge (Wort) über Start- und Endposition des Textes die geladenen Zusatzinformationen. Wenn Sie sich jetzt von einem Wort zum nächsten bewegen, fragt es diese Informationen über die Position des Cursors ab und zeigt sie u.a. in den Fenstern „Auflösung“ und „Lexikon“.

Es ergeben sich folgende Einschränkungen:

- Betrachten Sie den Text „*Rama geht ins Haus*“. In einem normalen Textfeld können Sie mit dem Cursor z.B. den Textabschnitt „*ama geht in*“ markieren und dann löschen, verschieben usw. Solche Aktionen sind in diesem Textfenster nicht erlaubt. Sobald Sie beginnen, einzelne Buchstaben zu markieren, ergänzt das Programm die Textmarkierung vom Anfang des ersten eingeschlossenen Wortes bis zum Ende des letzten. Damit wird die korrekte Verknüpfung von Lautgestalt und assoziierten Informationen sichergestellt.
- Sie können selbst nur die Laute des Nagari-Alphabets und <Shift+#> für den avagraha eingeben. Vers-, Strophen- und Satzenden sowie die Strophen-/Satzzählung werden vom Programm eingefügt.
- Besondere Erwähnung verdient die Funktion der TAB-Taste (s.u.). Mit ihr können Sie innerhalb der Sprachanalyse-Oberfläche zwischen den Auflösungen in der Auflösungsliste und dem Textfeld wechseln. Den aktuellen Eingabebereich (Text- oder Auflösungsfenster) erkennen Sie an der farbigen Umrahmung, die mit Betätigung der TAB-Taste hin- und herspringt.
- Die üblichen Tastaturbefehle Strg+X (markierten Text ausschneiden) und Strg+V (Text aus der Zwischenablage einfügen) sind in diesem Textfeld deaktiviert. Allerdings können Sie Textausschnitte per Strg+C in die Zwischenablage kopieren.

Im folgenden werden die erlaubten Tastatureingaben beschrieben:

Tastatureingabe	Beschreibung
Laute des Nagari-Alphabets,	Analog zu normalen Textfeldern.

avagraha,
Zahlzeichen
ENTER

Diese Eingabe wird nur behandelt, wenn sich der Cursor am Ende einer Lautfolge befindet. Wenn Sie einen Strophen- / Versumbruch mitten in einem Wort durchführen wollen, spalten Sie das Wort bitte zuerst an der gewünschten Stelle mit SPACE auf. Sie können zwei Arten von Satz- / Versumbrüchen einfügen:

- Wenn Sie nur die ENTER-Taste betätigen wird in metrischen Texten ein neuer Vers eingefügt. In Prosatexten – und Kommentaren – wird ein neuer Satz eingefügt.
- Wenn Sie Ctrl+ENTER betätigen, wird in metrischen Texten eine neue Strophe eingefügt. In Prosatexten hat diese Tastenkombination keine Auswirkungen.
- Fügen Sie mit den beschriebenen Möglichkeiten auch neue leere Verse, Strophen oder Sätze an eine Ebene an, indem Sie den Cursor am Ende des letzten Wortes eines Verses, ... positionieren und dann die gewünschte Befehlsfolge betätigen.

RETURN (Taste „<-
„)

Je nach Position des Cursors hat diese Taste fünf verschiedene Auswirkungen (Ähnlichkeiten zu normalen Textfeldern in Klammern hinter der Beschreibung):

- Innerhalb einer Lautfolge, die aus mehr als einem Buchstaben besteht: Buchstabe wird gelöscht, alle Auflösungen der Lautfolge werden ohne Nachfrage aus der Datenbank entfernt, weil die Lautgestalt verändert wurde.
- Besteht die Lautfolge aus einem Buchstaben, wird sie nach Nachfrage aus der Datenbank gelöscht (= ein Wort löschen).
- Sind mehrere Lautfolgen markiert, werden sie mitsamt ihrer Auflösungen nach Nachfrage aus der Datenbank gelöscht (= mehrere markierte Wörter löschen). Neben den Wörtern werden auch ihre Auflösungen gelöscht.
- Steht der Cursor am Anfang einer Lautfolge, wird sie mit der davorstehenden Lautfolge derselben Gliederungseinheit (z.B. Vers, Satz) verschmolzen. Alle Auflösungen werden aus der Datenbank gelöscht (= Freizeichen löschen).
- Steht der Cursor am Anfang einer Gliederungseinheit (Vers, Satz), wird diese Gliederungseinheit ohne Nachfrage mit der vorhergehenden verschmolzen, wenn eine solche vorhanden ist (= Löschen eines Zeilenumbruchs). D.h., wenn der Cursor z.B. an Beginn von Vers 2 positioniert ist, wird dieser Vers mit Vers 1 derselben Strophe verschmolzen. Wenn er dagegen vor Vers 1 von Strophe 1 einer Ebene, also ganz am Anfang der Ebene, positioniert ist, reagiert das Programm nicht auf Ihre Eingabe, d.h. dieser Vers wird nicht mit dem letzten des

	vorhergehenden Kapitels verschmolzen.
DEL (Taste „Entf“)	Funktioniert wie die RETURN-Taste, allerdings in entgegengesetzter Richtung.
SPACE (Freizeichen)	Mit dieser Taste spalten Sie eine Lautfolge auf (= Freizeichen). Alle zugeordneten Auflösungen werden ohne Nachfrage aus der Datenbank gelöscht. Befindet sich der Cursor am Ende eines Wortes, erstellen Sie mit der SPACE-Taste einen neuen Texteintrag (= neues Wort schreiben). Wenn Sie die SPACE-Taste am Wortende betätigt haben und den Cursor weiterbewegen wollen, ohne ein neues Wort zu schreiben, setzt das Programm den Cursor auf den Freiraum zurück.
TAB (Tabulator; „Zeile einrücken“)	Mit dieser Taste wird der Eingabefocus vom Textfeld auf die Liste der Auflösungen im Auflösungs Fenster gesetzt. Sie können jetzt über die Pfeiltasten nach links und rechts in der Auflöungsliste navigieren. Wenn Sie sich im Auflösungs Fenster befinden und die TAB-Taste drücken, gelangen Sie wieder zurück in das Textfeld.

Die Funktionen der Tastatur werden durch folgendes Kontextmenü (Aufruf über rechte Maustaste) erweitert:

Befehl	Shortcut	Beschreibung
Vers- -> Strophenumbruch		Wandelt einen Vers- in einen Strophenumbruch um (d.h. verwandelt zeilenbeendes / in //), wenn der Cursor in einem versbeendenden Wort steht.
Strophen- -> Versumbruch		Wandelt einen Strophen- in einen Versumbruch um (d.h. verwandelt zeilenbeendes // in /), wenn der Cursor in einem strophenbeendenden Wort steht.
Lücke ...		Öffnet den Dialog „Textlücke erstellen / bearbeiten“, mit dem Sie Auftreten und Umfang von Textlücken in Originalmanuskripten beschreiben können. Wenn sich der Cursor bei Aufruf dieses Befehls in einer Lücke befindet, können Sie deren Eigenschaften bearbeiten.
Neue Lesart	F3	Öffnet den Dialog „Lesart erstellen / bearbeiten“, in dem Sie eine neue Lesart im aktuellen Text erstellen können. Wenn Sie keinen Text markiert haben, wird die Lesart der Lautfolge zugeordnet, in der sich momentan der Cursor befindet. Wenn Sie mehrere Lautfolgen markiert haben, wird die Lesart diesen Lautfolgen kollektiv zugewiesen. Sie können mehrere Lesarten für eine Lautfolge eintragen.
Lesart bearbeiten ...	Shift+F3	Wenn sich der Cursor im Text einer Lesart befindet (d.h. nicht im Haupttext), öffnet sich der Dialog „Lesart erstellen / bearbeiten“, in dem Sie die Informationen zu der Lesart bearbeiten können.
In Lexikon suchen ...	F6	Dieser Befehl öffnet die Lexikonoberfläche. Als Voreinstellung für das Suchwort ist die Lautfolge

geladen, in der sich der Cursor befindet.

Probleme bei der Sprachanalyse

In manchen Fällen werden Wörter nicht in der gewünschten Weise oder gar nicht analysiert. In den meisten Fällen handelt es sich dabei um eines der folgenden Probleme:

- **Orthografie:** Überprüfen Sie die Rechtschreibung des Wortes! Falsch geschriebene Wörter werden vom Programm nicht korrekt analysiert. Ein häufiges Problem ist der Ersatz des anusvara durch den Klassennasal. Schlagen Sie in diesem Fall die korrekte Form des Lemmas im Lexikon nach, und verändern Sie dementsprechend den Text.
- **Unbekanntes Lemma:** Stellen Sie sicher, dass alle Wörter, die in einem Kompositum vorkommen, auch wirklich im Lexikon enthalten sind. Wenn eines der Wörter nicht gefunden werden kann, wird das gesamte Kompositum nicht analysiert. Wenn das gesuchte Wort im Lexikon enthalten ist, sollten Sie sicherstellen, dass dem Wort auch die gesuchte grammatikalische Kategorie zugeordnet ist. Wenn nicht, weisen Sie einer Nominalform die gesuchte Kategorie im Dialog „Lemma erstellen / bearbeiten, Register Nominalformen“ zu! Wenn es sich bei dem Wort um ein Verb handelt, sollten Sie im Dialog „Verbalformen“ überprüfen, ob dem Verb die gesuchte Form zugeordnet worden ist. – Ist das gesuchte Wort noch nicht im Lexikon enthalten, erstellen Sie es neu (Dialog „Lemma erstellen / bearbeiten“), und wiederholen Sie die Sprachanalyse!
- **Wortschatz:** Wenn die Grammatikanalyse mit aktiviertem Wortschatz durchgeführt wurde, können neue Wörter zugunsten bekannter unterdrückt werden. Deaktivieren Sie in diesem Fall den Wortschatz, und analysieren Sie den Text erneut.
- **Programmabsturz:** Wenn Sie ein Lemma in das Lexikon eingetragen haben und danach – was nicht geschehen möge – ein Programmabsturz stattgefunden hat, ist dieses neue Lemma (oder auch eine neue Kategorie eines bekannten Lemmas) zwar in der Datenbank, aber **NICHT** im Spezial-Wörterbuch zur Sprachanalyse gespeichert. Gehen Sie in diesem Fall wie folgt vor:
 - Schliessen Sie das Programm.
 - Suchen Sie (z.B. mit dem ®Windows-®Explorer©) eine Datei namens „Sanskritlexikon.cba“, die sich normalerweise im Ordner des Programms befinden sollte.
 - **Löschen** Sie diese Datei (d.h. die Datei „Sanskritlexikon.cba“).
 - Starten Sie das Programm von neuem!! Der Programmstart nimmt einmalig eine deutlich längere Zeit in Anspruch; allerdings sollte danach das Lexikon wieder problemlos funktionieren.
- **Wortlänge:** Sehr lange Komposita mit vielen Sandhi-Möglichkeiten werden aus Zeit- und Speichergründen nicht analysiert. Aktivieren Sie zunächst den Wortschatz, der

den Umfang des Lexikons auf bisher schon angetroffene Lemmata beschränkt. Wenn das Wort immer noch nicht analysiert werden kann, sollten Sie die Auflösung manuell erstellen oder es unaufgelöst lassen.

Im-/Export: Texte einfügen und exportieren

Text importieren

Neben der manuellen Eingabe von neuem Text haben Sie die Möglichkeit, bestehende Textdateien in einer Reihe von vordefinierten Formaten in die Datenbank einzulesen. Sobald eine solche Textdatei korrekt importiert ist, kann sie mithilfe des Grammatikmoduls lexikalisch und morphologisch analysiert werden.

ACHTUNG: Wenn ein Text, den Sie einlesen wollen, aus mehreren Kapiteln besteht, müssen Sie ihn vor dem Einlesen in einem externen Text-Editor in die einzelnen Kapitel aufteilen. Wenn Sie den gesamten Text auf einmal importieren, ergeben sich unvorhersehbare Resultate bei der Vers- und Strophennumerierung und eine Reihe von unangenehmen Fehlermeldungen.

Die nötigen Schritte werden im folgenden beschrieben.

Schritt 1: Vorbereitung des Originaltextes in einem externen Texteditor

- Teilen Sie den Originaltext in Kapitel auf.
- Entfernen Sie unnötige Freizeilen im Originaltext.
- Entfernen Sie Formatierungshinweise im Originaltext. Der Text sollte nach dieser Vorbehandlung wie das folgende Beispiel aussehen:

```
atha vakṣye mantrakośaṃ yad uktaṃ bhūtaḍāmare /
kramād bījavidhānaṃ ca śrutyaṛthajñānasambhavam //
```

D.h.:

- Sämtliche numerischen Formatierungsangaben – v.a. den Versen vorgestellte Markierungen - sind entfernt. **ACHTUNG:** Verwenden Sie für solche erweiterten Ersetzungen z.B. die Befehle “reguläre Ausdrücke verwenden”, “mit Mustervergleich” o.ä. in ihrem Texteditor.
- Zeilen- und Strophenenden sind durch eindeutig erkennbare Identifizierer markiert. Diese Identifizierungen können frei gewählt werden (ein Strophenende kann z.B. durch **, ein Versende durch * markiert sein). Die Identifizierungen müssen innerhalb des gesamten Textes konsistent sein. Sie dürfen keine numerischen Angaben enthalten (d.h. Identifizierer der Form ... // 5 // sind nicht erlaubt).
- Sämtliche Sandhimarkierungen sind aus dem Text entfernt worden.
- Der Text liegt in einem der folgenden Transkriptionssysteme vor: der datenbank-internen Transkription, CSX, Kyoto-Transkription.

Sobald der Text diese Anforderungen erfüllt, kann er in die Datenbank eingelesen werden.

Schritt 2: Den Import-Dialog öffnen

Öffnen Sie mit dem Menübefehl “Sprachanalyse” -> “Text importieren” den Dialog “Text einlesen”. In diesem Dialog müssen Sie Angaben zur Textherkunft, zur Umwandlung der Transkription (falls nötig) und der Gliederungsmethode des Textes machen.

Schritt 3: Textdatei auswählen

Betätigen Sie den Knopf “Suchen ...” oben im Dialog. Ein Windows-typischer Datei-Dialog wird geöffnet, in dem Sie die Textdatei auswählen, die Sie in die Datenbank einlesen wollen. Nach Schliessen des Dialogs wird eine Textvorschau rechts im Fenster geladen.

Schritt 4: Den Zieltext suchen

In diesem Schritt müssen Sie die bibliografischen Details zum einzulesenden Text angeben. Die Befehle zum Erstellen und Bearbeiten von Textvorlagen finden sich in einer Reihe wunderschöner kleiner Knöpfe unter dem grossen Baumdiagramm in der Mitte des Fenster. Für Informationen über die Funktion der einzelnen Tasten führen Sie den Cursor über eine der Tasten und lassen Sie ihn für einen Augenblick an dieser Stelle stehen.

Beispiel 1: Sie wollen einen kurzen Text vom Namen xy, der noch nicht in der Datenbank enthalten ist, in die Datenbank einlesen. Gehen Sie wie folgt vor:

- Betätigen Sie die Taste “Neuen Text auswählen”. Der Dialog “Textvorlage auswählen/erstellen” erscheint. Beim ersten Erscheinen dieses Dialogs werden Sie gefragt, ob Sie einen neuen Text erstellen wollen. Bestätigen Sie mit “Ja” (wenn Sie aus Versehen auf “Nein” gedrückt haben, können Sie einen neuen Text mit der Taste “Neue Textvorlage” erstellen).
- Füllen Sie jetzt die entsprechenden Felder mit den bibliografischen Angaben.
- Schliessen Sie den Dialog mit “Übernehmen”.
- Wählen Sie jetzt die erste Taste (“Neue Ebene (untergeordnet)” == neues Kapitel). Geben Sie im folgenden Dialog der Ebene einen Namen. Die bibliografischen Angaben sind jetzt vollständig.

Beispiel 2: Sie haben das erste Kapitel eines Textes eingelesen und wollen jetzt das nächste einlesen. Gehen Sie wie folgt vor:

- Betätigen Sie die Taste “Neuen Text auswählen”. Wählen Sie bei Erscheinen des Dialogs “Textvorlage auswählen/erstellen” die Option “Nein” (Sie wollen keinen neuen Text erstellen, sondern einen vorhandenen auswählen!).
- Wählen Sie aus der Auswahlliste “Suchen” links unten im Dialog den gesuchten Text aus.
- Schliessen Sie den Dialog mit “Übernehmen”.
- Im Baumdiagramm sollte jetzt mindestens eine Ebene (= Kapitel) erscheinen. Markieren Sie diese Ebene mit dem Cursor.

- Betätigen Sie jetzt die zweite Taste von links (“Neue Ebene (gleichgeordnet)”), und geben Sie dem neuen Kapitel einen Namen.

Schritt 5: Umwandeln der Transkription

Wählen Sie aus der Liste die im Text benutzte Transkriptionsmethode aus. Der Vorschautext verändert sich gemäss Ihren Angaben. **ACHTUNG:** Der Text im Vorschauenfenster zeigt die Gestalt, in der der Text in der Datenbank aufgenommen wird. Bitte überprüfen Sie den von Ihnen vorbereiteten Text auf nicht-singuläre Fehler/Inkonsistenzen in der Transkription!

Schritt 6: Formatierung festlegen

Wählen Sie eine der Optionen aus der Formatierungsliste. Wenn Sie die Option “benutzerdefiniert” auswählen, müssen Sie in den Textfeldern unter der Liste die Markierungen für Strophen- (z.B. //) und Versenden (z.B. /) angeben.

Der Text kann jetzt mit der Schaltfläche “Einlesen” in die Datenbank eingefügt und danach z.B. lexikalisch analysiert werden.

Text exportieren

Den im Hauptfenster dargestellten Text können Sie mit dem Menübefehl “Sprachanalyse” -> “Text in die Zwischenablage” formatiert in die Zwischenablage und von dort in ein externes Textprogramm einfügen. Achten Sie darauf, dass in diesem Programm die korrekten Fonts geladen sind.

Wenn Sie einen Text vollständig auslesen wollen, wählen Sie die Funktion “Weitere Funktionen” -> “Gesamten Text exportieren”. Im folgenden Dialog können Sie die gewünschte Zieltranskription angeben. Nach dem Export, der einige Minuten in Anspruch nehmen kann, wird der Text als .txt-Datei auf Ihrem Datenträger gespeichert.

Verbalformen erstellen/bearbeiten

In diesem Dialog erstellen und verwalten Sie Verbalparadigmata. Das Programm kennt die wichtigsten Bilderegeln für Verbalformen, so dass Sie neue Paradigmata meist per Knopfdruck einfügen können. Verbalformen, die in den Texten belegt sind, werden unterstrichen; die Fundstellen können angezeigt werden. Die Benutzerführung des Dialogs ist etwas gewöhnungsbedürftig!!

Der Dialog teilt sich in drei Hauptbereiche:

- Links finden Sie drei Listen, auf die die Verbalstämme je nach Bildweise verteilt sind. Die wichtigste ist das oberste Auswahlfeld „Basisverben“. In diesem Feld finden

Sie alle nicht-zusammengesetzten Verbalwurzeln. Wenn Sie einen Eintrag, ein „Basisverb“ aus dieser Liste wählen, werden in der Liste „Sekundäre Konjugation“ alle vom Basisverb abgeleiteten sekundären Verbalwurzeln aufgeführt. In der Liste „Zusammengesetzte Verben“ sind die durch Präfigierung entstandenen Ableitungen des Basisverbs zu finden. Gleichzeitig wird das Basisverb zum aktuellen Verb gemacht.

- Links unten befindet sich ein Textrahmen („Aktuelles Verb“), der alle Informationen über das aktuell ausgewählte Verb umschließt. Das aktuelle Verb haben Sie entweder aus der Auswahlliste „Basisverben“, aus der Liste „Sekundäre Konjugation“ oder der Liste „Zusammengesetzte Verben“ ausgewählt. Zunächst sind links ein kurzer Text, der die Wurzel beschreibt, und der Bindevokaltyp abgedruckt. Darauf folgt die Liste „Zugeordnete Präsensklassen“. Diese Liste umfasst alle Präsensklassen, die dem aktuellen Verb zugeordnet wurden. Wenn die von Ihnen gewünschte Klasse nicht in der Liste aufgeführt ist, betätigen Sie den Knopf „Verb bearbeiten“, der den Dialog „Lemma erstellen/bearbeiten“ öffnet. Hier können Sie dem Verb neue Präsensklassen zuordnen. Wenn Sie eine Präsensklasse aus der Liste auswählen, wird nach kurzem Warten die nächste Liste, „Behandelte Tempora / Modi“ gefüllt. Diese Liste markiert alle Tempora / Modi für die aktuelle Präsensklasse, für die schon Formen vorhanden sind, durch einen Haken neben dem Namen des Tempus.
- Mit der Aktualisierung der Tempus-Liste wird das zentrale Fenster des Dialogs gefüllt. Hier sind alle eingegebenen Formen und Nebenformen des aktuellen Verbs aufgezählt.

Im folgenden finden Sie Anleitungen für die wichtigsten Aktionen.

Das „aktuelle Verb“ festlegen:

Diese Aktion unterscheidet zwischen Basisverben, Verben der Sekundärkonjugation und zusammengesetzten Verben.

- *Basisverb*: Wählen Sie das Verb aus der Auswahlliste aus.
- *Sekundäre Konjugation*: Wählen Sie das Basisverb aus der Auswahlliste und dann das sekundäre Verb aus der Liste „Sekundäre Konjugation“.
- *Zusammengesetzte Verben*: Wenn das zusammengesetzte Verb mit Präfix von einem Basisverb abgeleitet ist, wählen Sie erst das Basisverb und dann aus der Liste „Zusammengesetzte Verben“ das gewünschte Verb; ist es von einem sekundären Verb abgeleitet, wählen Sie erst das sekundäre und dann das gewünschte Verb.

Ein Tempus / Modus eines Verbs neu bilden

- Machen Sie das Verb zum aktuellen Verb (s.o.).
- Wählen Sie die gewünschte Präsensklasse des aktuellen Verbs aus. Wenn in der Liste „Tempus / Modus“ neben dem gewünschten Tempus noch kein Haken ist, aktivieren Sie den Haken mit der Maus. Das Paradigma wird erstellt und im

Hauptfenster abgedruckt. **ACHTUNG:** Überprüfen Sie immer die Ergebnisse dieser Prozedur! Gerade bei untypischen Formen und im Perfekt schleichen sich gern Fehler ein!

- Für einige Tempora / Modi ist die automatische Formenbildung nicht implementiert. In diesem Fall aktivieren Sie zuerst den Haken neben dem Tempus. Dann klicken Sie doppelt auf das Paradigma-Rudiment, wodurch sich ein Dialog öffnet, in dem Sie die Formen per Hand eintragen können.

Mehrere Tempora / Modi eines abgeleiteten Verbs bilden

Diese Vorgehensweise wird empfohlen, wenn Sie das Paradigma zusammengesetzter Verben erstellen wollen, deren Basisverben schon behandelt wurden (z.B.: das Paradigma von adhi*gam, wenn die Formen für *gam schon vorhanden sind).

- Leiten Sie das Verb korrekt von seinem Basisverb ab, d.h. machen Sie korrekte Angaben im Dialog „Lemma erstellen/bearbeiten“ (s. S. 12).
- Wenn das Verb, von dem das aktuelle abgeleitet ist, schon Formen für seine Präsensklassen besitzt, aktivieren Sie die Option „automatisch aktualisieren“. Wenn Sie jetzt eine Präsensklasse des aktuellen, abgeleiteten Verbs markieren, erscheint der Dialog „*Formenergänzung*“. In diesem Dialog sind alle Tempora des Basisverbs aufgelistet, die im Paradigma des aktuellen Verbs noch nicht behandelt wurden (Mehrfachauswahl per SHIFT / CTRL möglich). Markieren Sie in der Liste die Tempora, die automatisch gebildet werden sollen. Schliessen Sie den Dialog mit „Übernehmen“.
- Das Programm bildet jetzt auf Basis der schon eingegebenen Formen des Basisverbs die Formen des zusammengesetzten Verbs.
- Überprüfen Sie nach dem Erstellen unbedingt die Binfuge zwischen Präfixen und Verbalstamm, da das Programm keinen wortinternen Sandhi durchführt. Korrigieren Sie eventuelle Fehler nach Doppelklick auf das Paradigma im Dialog „Tempus-Paradigma“ (für finite Formen) bzw. „Verbalnomina“ (für infinite Formen)!

Ein Tempus / Modus eines Verbs löschen

- Aktivieren Sie das gewünschte Verb (s.o.).
- Wählen Sie die gewünschte Präsensklasse. Wenn in der Liste „Tempus / Modus“ neben dem Tempus, das Sie löschen wollen, ein Haken zu sehen ist, deaktivieren Sie diesen Haken. Nach einer Nachfrage wird das gesamte Tempusparadigma gelöscht. **ACHTUNG:** Sie löschen mit dem Paradigma auch alle möglichen Fundstellen der gelöschten Formen innerhalb der Texte, die in der Datenbank verwaltet werden.

Eine Einzelform bearbeiten

- Aktivieren Sie das gewünschte Verb (s.o.).
- Laden Sie die gewünschte Präsensklasse.

- Suchen Sie die Form im Hauptfenster, und klicken Sie doppelt auf den Text der Form. Der Dialog „Tempus-Paradigma“ wird geöffnet. Hier können Sie die Form nach Herzenslust bearbeiten.

Eine Einzelform löschen

- Laden Sie das gewünschte Paradigma.
- Klicken Sie auf die Form, die Sie löschen wollen, und bestätigen Sie die ENTF-Taste. Die Form wird gelöscht. **ACHTUNG:** Sie löschen mit der Form alle Verknüpfungen, die eventuell zu gespeicherten Texten bestehen.
- Alternativ können Sie das gesamte Tempusparadigma per Doppelklick in den Dialog „Tempus-Paradigma“ laden (vgl. voriger Schritt) und die Form an dieser Stelle löschen.

Eine neue Einzelform erstellen

- Laden Sie das gewünschte Paradigma.
- Laden Sie den Dialog „Tempus-Paradigma“ per Doppelklick auf das Paradigma, und erstellen Sie hier die gewünschte Form.

Nominale Sonderformen erstellen/bearbeiten

Mit diesem Dialog bearbeiten Sie Nomina – darunter fallen auch Pronomen und Zahlwörter -, die abweichend von den normalen Bildungsregeln erstellt werden. Geben Sie entweder das gesamte Paradigma ein oder die Formen, die belegt sind oder Ihnen interessant vorkommen.

Um ein **neues Paradigma zu erstellen** gehen Sie wie folgt vor:

- Geben Sie im Lexikon-Suchfeld oben links das Wort ein. Betätigen die ENTER-Taste. Wählen Sie dann den Treffer aus der Liste per Doppelklick aus. Das Lemma erscheint im Textfeld „Aktuelles Wort“.
- Wenn das Wort mehrere Genera besitzt, wählen Sie das gewünschte aus der Liste aus.
- Tragen Sie jetzt die Formen in die entsprechenden Felder ein.
- Wenn Sie für eine Form Neben- oder vedische Formen erstellen wollen, setzen Sie den Cursor in das Feld der entsprechenden Form und betätigen Sie dann den Knopf „Neue Nebenform“.
- Wiederholen Sie die Schritte für andere Lemmata oder weitere Genera desselben Lemmas.

Beim Wechsel zu einem anderen Genus, beim Aktivieren eines neuen Lemmas und beim Schliessen des Dialogs fragt das Programm nach, ob eventuelle Änderungen gespeichert werden sollen.

Wenn einem Lemma schon Sonderformen zugeordnet worden sind, erscheint es in der Liste „Wörter mit Sonderformen“ oben rechts im Dialog. Aktivieren Sie das Lemma durch Doppelklick auf den Eintrag.

Sie löschen Formen, indem Sie ihren Text löschen (d.h. z.B. Text markieren und DEL drücken; oder auf eine andere Weise).