

Altenglisch mit Shoebox

Einleitung Im Folgenden soll ein neues Projekt begonnen werden, in dem es darum geht, Daten aus der spätaltenglischen Überlieferung morphologisch zu analysieren und zu interlinearisieren. Ein besonderes Merkmal wird dabei sein, dass der Ausgangstext in einer besonderen Schriftart für das Altenglische (Junius Modern) verfasst wird, in der z.B. Vokallängen (ǣ ā ē ī ō ū) und diakritische Zeichen für die Palatalisierung (č und ġ) angezeigt werden können, und dass der Analyse-Output auch eine phonetische Umschrift enthalten soll:

þā	cām	þær	reġen	and	micel	flōd,	and	þær	blēowon	windas
þā	cum	þær	reġen	and	micel	flōd	and	þær	blēow	-on wind -as
da	komm.Prät	dort	Regen	und	groß	Flut	und	dort	blas.Prät -pl	wind -Nom.Pl
Adv	V	Pron	N	Konj	A	N	Konj	Pron	V	-suff N suff
θa:	ko:m	θæ:r	reġen	and	mitʃel	flo:d	and	θæ:r	ble:əw	-on wind -as

Daraus ergeben sich eine Reihe neuer Probleme, die in den folgenden Abschnitten an geeigneter Stelle diskutiert werden sollen.

Was ist zu tun?

- Wir müssen für unsere Aufgabe zunächst einige Vorarbeiten erledigen. Zunächst muss eventuell das **Tastaturlayout** für die altenglische Schrift für das persönliche Konto aktiviert werden.
- Damit die Vorarbeiten nicht zu viel Zeit in Anspruch nehmen, wurden einige Konfigurationsdateien (Datenbanktypen für das Lexikon und den Interlineartext) bereits erstellt. Diese müssen von den Internetseiten geholt und in Ihr Shoebox-Verzeichnis kopiert werden.
- Die altenglischen Textdaten müssen erfasst und für die Interlinearisierung aufbereitet werden.
- Damit möglichst schnell mit der Interlinearisierung begonnen werden kann, sollen zunächst einige Paradigmen aus der altenglischen Grammatik in das Lexikon aufgenommen werden.

Aktivieren des Tastaturlayouts

Überprüfen Sie zunächst, ob das Tastaturlayout für Altenglisch für Sie schon verfügbar ist. Klicken Sie dazu auf das blau unterlegte **DE** in der Taskleiste. Wenn Sie dort das Tastaturlayout nicht finden sollten, müssen Sie es anhand des separaten Textes mit der Überschrift Tastaturbelegung für Altenglisch aktivieren. Sie brauchen dies nur einmal zu tun. Die neue Einstellung wird Bestandteil ihres persönlichen Profils.

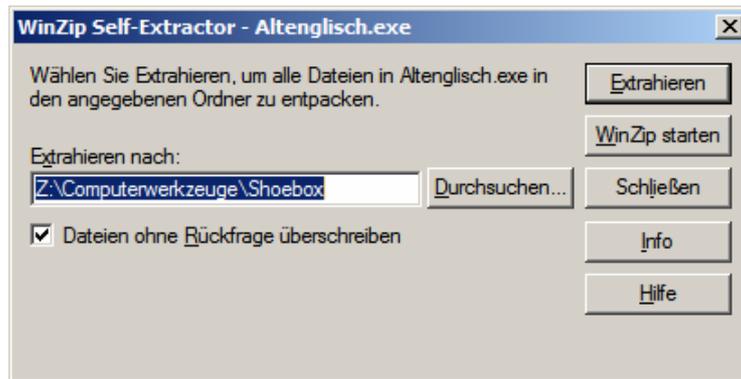


Einrichtung der Verzeichnisstruktur für das Altenglischprojekt

Auf den Internetseiten zur Veranstaltung finden Sie eine Verknüpfung zu einer Datei mit der Bezeichnung **Altenglisch.exe**. Es handelt sich dabei um ein selbstentpackendes Dateiarchiv, das eine Reihe von für das Altenglischprojekt notwendigen Dateien enthält. Speichern Sie dieses Archiv in Ihrem Verzeichnis **Z:\Computerwerkzeuge\Shoebox**.

- Es ist wichtig, dass Sie diese Verzeichnisstruktur haben. Überprüfen Sie dies und nehmen Sie ggf. Korrekturen vor. Im Zweifelsfall fragen Sie den Veranstalter.

- Öffnen Sie den Browser und navigieren Sie zur Startseite der Veranstaltungen (www.fb10.uni-bremen.de/linguistik/khwagner), wählen Sie Morphologie und dort den Menüpunkt Morphologie mit Shoebox. In der Karteikarte mit der Bezeichnung Altenglisch mit Shoebox klicken Sie mit der rechten Maustaste auf das Winzip Icon  und wählen Sie Ziel speichern unter und speichern Sie die Datei im Verzeichnis **Z:\Computerwerkzeuge\Shoebox**
- Gehen Sie jetzt zu diesem Verzeichnis. Sie sollten dort die Datei **Altenglisch.exe** finden. Öffnen Sie die Datei durch Doppelklick. Es öffnet sich folgendes Dialogfenster. Klicken Sie auf den Schalter Extrahieren, um das Archiv zu entpacken.



Wenn Sie das letzte Mal alles richtig gemacht haben, müßte jetzt in Ihrem Shoeboxverzeichnis ein Unterverzeichnis Altenglisch vorhanden sein, das eine Reihe von für das Altenglischprojekt erforderlichen Dateien enthält. Außerdem wurde im Shoeboxverzeichnis selbst die Datei **textprep.cct** abgespeichert, die für die Aufbereitung von Textdateien für die Interlinearisierung wichtig ist.

Erfassung der altenglischen Textdaten

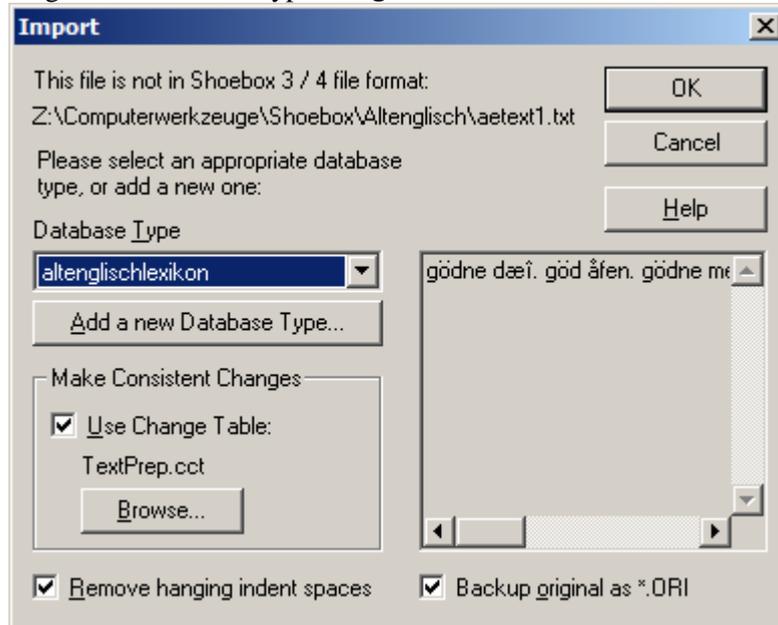
Als nächstes sollen Sie eine Textdatei mit den altenglischen Daten auf dem Aufgabenblatt erstellen. Dabei ist nur der altenglische Text zu übernehmen. Abweichend von der bisherigen Praxis sollen Sie für die Texterfassung das Programm Word benutzen. Gehen Sie dabei folgendermaßen vor:

- Starten Sie das Programm **Word**.
- Aktivieren Sie in der Taskleiste das altenglische Tastaturlayout.
- Stellen Sie die Anzeigegröße auf Seitenbreite oder wählen Sie einen prozentualen Vergrößerungsgrad.
- Schreiben Sie den Text exakt wie in der Vorlage. Hinweise über die Zuordnung der altenglischen Sonderzeichen zu den Tastaturcodes finden Sie auf der letzten Seite des Textes über das altenglische Tastaturlayout.
- Wenn Sie mit der Erfassung des Textes fertig sind, dürfen Sie ihn nicht als normales Worddokument (Erweiterung *.doc) abspeichern. Wählen Sie Datei Speichern unter ... und als Dateityp Nur Text (Erweiterung *.txt). Speichern Sie die Datei unter dem Namen **aetext1.txt** im Verzeichnis **Altenglisch** ab.

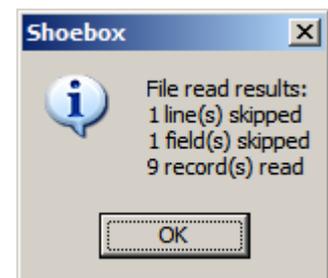
Text für Interlinearisierung vorbereiten

Der Originaltext, der interlinearisiert werden soll, muß für diesen Zweck aufbereitet werden. Er muß z.B. in Datensätze zerlegt werden, deren Felder aus den Sätzen bzw. Teilsätzen, der morphologischen Analyse, der Glossierung und einer freien Übersetzung bestehen. Gegenenfalls können noch andere Felder hinzukommen, z.B. ein Kommentarfeld. Bei dieser Aufgabe hilft uns allerdings Shoebox.

- Starten Sie das vordefinierte Altenglischprojekt, indem Sie die Projektdatei **altenglisch.prj** im Verzeichnis Altenglisch doppelt klicken. Das Shoebox-Hauptfenster ist leer, weil bisher weder eine Lexikon angelegt worden ist, noch ein Text importiert wurde.
- Wählen Sie *File Open* und laden Sie den Text **aetext1.txt** es erscheint das folgende Dialogfenster. Es teilt beispielsweise mit, daß die Datei nicht in einem Shoebox-Format ist und fordert dazu auf, einen geeigneten Datenbanktyp zuzuordnen. Angeboten wird der Typ **altenglischlexikon**, was natürlich falsch ist.



- Es ist allerdings schon ein Datenbanktyp für die Interlinearisierung von altenglischen Texten namens **altenglischtext** definiert worden. Wählen Sie diesen aus. Aktivieren Sie unbedingt auch das Feld **Use Change Table**. Die dafür erforderliche Datei **TextPrep.cct** müsste ja jetzt vorhanden sein. Klicken Sie anschließend auf OK.
- Nach dem erfolgreichen Import der Textdatei erhalten Sie eine Mitteilung, die u.a. anzeigt, wieviele Datensätze bei der Konversion erzeugt worden sind. Klicken Sie hier wiederum auf den OK-Schalter.



aetext1.txt	
\ref	001
\t	gödne dæg.
\f	

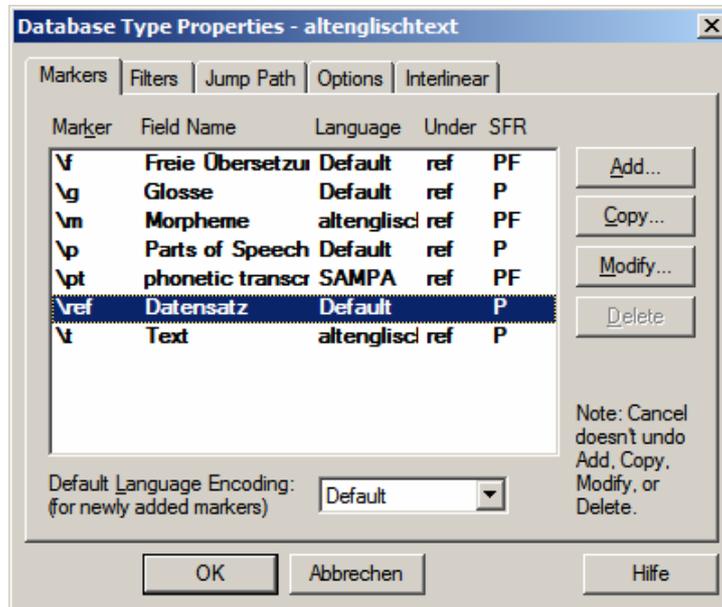
Zum Abschluss wird nun der erste Datensatz des importierten Textes angezeigt. Er enthält zunächst 3 Datenfelder, ein Feld mit der Markierung (marker) **\ref** mit einer Referenznummer, ein Feld mit der Markierung **\t** für den Ausgangstext, und ein Feld **\f**, das für eine freie (idiomatische) Übersetzung gedacht ist. Wichtig ist, dass der Ausgangstext mit dem Zeichensatz **Junius Modern** angezeigt wird. Wir werden bei der Interlinearisierung feststellen, dass neben mehreren anderen ein Feld **\pt** für die phonetische Transkription eingefügt

wird, das eine IPA-Zeichensatz verwendet.

Eigenschaften der Interlinearisierung kontrollieren

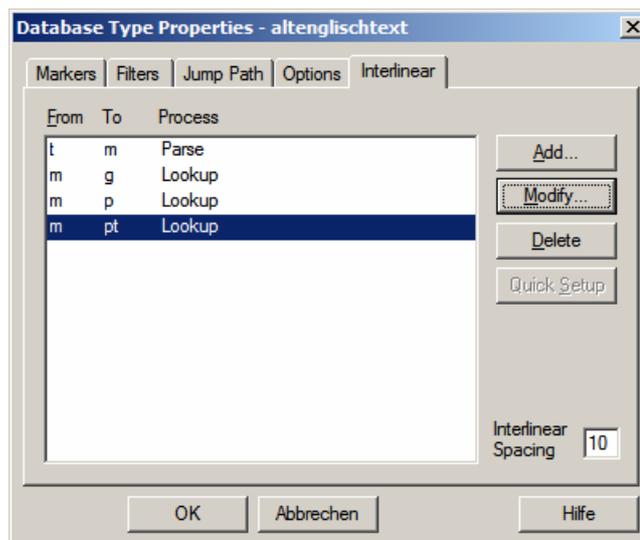
Zum besseren Verständnis ist es sicher nicht verkehrt, sich die Eigenschaften des Datenbanktyps, der dem soeben importierten Text zugeordnet ist, etwas genauer anzusehen, damit deutlich wird, was sich bei der Interlinearisierung abspielt.

- Aktivieren Sie das Fenster mit der Überschrift **aetext1.txt**.
- Wählen Sie im Menü Databases den Punkt Properties ... Das folgende Fenster zeigt zunächst die verwendeten Markierungen (Markers).

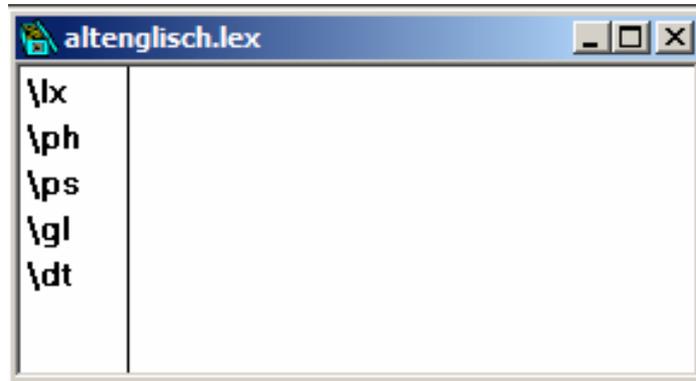


Gegenüber dem Bisherigen fällt auf, dass eine Markierung **\pt** mit dem Namen phonetic transcription hinzugekommen ist, und zwar mit einer eigenen Sprachkodierung **SAMPA** (dazu später mehr). Den Markierungen **\m** (Morpheme) und **\t** (Text) ist eine Sprachkodierung für Altenglisch zugeordnet.

- Aktivieren Sie jetzt die Karteikarte Interlinear. Auch hier ist etwas hinzugekommen, nämlich die Zuordnung von Morphemen (m) zur phonetischen Transkription, was durch Nachschlagen (Lookup) im Wörterbuch geschieht.



Altenglisch-Lexikon und Interlinearisierung Für die Interlinearisierung unserer Textdaten wurde bereits ein Lexikon namens **altenglisch.lex** angelegt, das allerdings noch leer ist. Öffnen Sie mit File Open dieses Lexikon. Wie sie sehen können enthält es neben den bereits bekannten Feldern ein weiteres Feld für die phonemische Transkription (**\ph**)



- Machen Sie zunächst Lexikoneinträge für die Paradigmen auf der Rückseite des Datenblattes (Altenglische Konversation).
- Die phonemische Umschrift wird im SAMPA-Kode eingegeben. Erläuterungen dazu finden Sie auf dem Blatt **SAMPA für Altenglisch**.
- Beginnen Sie dann mit der Interlinearisierung der altenglischen Textbeispiele und ergänzen Sie erforderlichenfalls die Lexikoneinträge. Im Zweifelsfall brauchen Sie zunächst keine tiefgehende Analyse zu machen. Im Vordergrund steht zunächst die phonemische Transkription.