

Das Esperanto-Projekt

In diesem Abschnitt sollen Sie eine kleine lexikalische Datenbank mit Esperanto-Daten erstellen. Im Unterschied zum Testprojekt oder dem Lateinprojekt kommt hier der Umstand hinzu, dass das **Esperanto** über eine Reihe von **Sonderzeichen** verfügt, die Shoebox bekannt gemacht werden müssen, damit es die Daten interpretieren kann. Schritt für Schritt werden Sie die folgenden Aufgaben ausführen müssen:

1. Erstellen eines neuen Projekts.
2. Erstellen eines Datenbank-Typs.
3. Definition der notwendigen Sprachkodierungen (*Language Encodings*)
4. Erstellen einer neuen Datenbank-Datei für das Lexikon. Dies beinhaltet die folgende Schritte:
 - 4.1. Erstellen des ersten Datensatzes
 - 4.2. Verknüpfung der Datensatz-Markierungen mit den Sprachkodierungen
 - 4.3. Erstellen eines Datensatz-Musters (engl. *Template*)

Wie Sie sehen, sind die Punkte 3 und 4.2 von besonderem Interesse, da es in ihnen um die Sonderzeichen geht.

Bereitstellung des Esperanto Zeichensatzes

Das Schriftsystem von Esperanto verwendet eine Reihe von Sonderzeichen insbesondere für die Kodierung von Sibilanten (/ʃ, ʒ, tʃ, dʒ/), des Phonems /x/, sowie für den Gleitlaut /w/.

Esperanto verwendet folgende Vokalbuchstaben (jeweils mit Groß- und Kleinschreibung), wobei die Aussprache der IPA-Symbole entspricht:

A, a – E, e – I, i – O, o – U, u

Esperanto hat 23 Konsonanten (genauer 21 Konsonanten und zwei Gleitlaute oder Halbvokale: j, ŭ), die wie folgt kodiert werden:

B, b – C, c /ts/ – Ĉ, ĉ /tʃ/ – D, d – F, f – G, g – Ĝ, ĝ /dʒ/ – H, h – Ĥ, ĥ /x/ – J, j – Ĵ, ĵ /ʒ/ – K, k – L, l – M, m – N, n – P, p – R, r – S, s – Ŝ, ŝ /ʃ/ – T, t – Ŭ, ŭ /w/ – V, v – Z, z /z/.

Die Buchstaben Q, q, W, w und X, x kommen nicht vor.

Alle Sonderzeichen sind im Standard-Unicode Zeichensatz von Windows NT zwar vorhanden, aber leider versteht Shoebox kein Unicode. Es wäre daher eigentlich sinnvoll, einen eigenen Zeichensatz zu kreieren (es gibt auch solche Zeichensätze). Wir werden statt dessen jedoch eine Konvention einführen, die im Internet Standard geworden ist, und alle Sonderzeichen durch Hinzufügen von *X* bzw. *x* markieren, das ja nicht selbst Teil des Esperanto-Alphabets ist. Ĉ, ĉ – Ĝ, ĝ – Ĥ, ĥ – Ĵ, ĵ – Ŝ, ŝ – Ŭ, ŭ werden also als Cx, cx – Gx, gx – Hx, hx – Jx, jx – Sx, sx – Ux, ux kodiert. Wir schreiben also für *ĉeko* *cxeko*, für *ĝardeno* *gxardeno* und für *Ĥoro* *Hxoro*.

Erstellung eines neuen Projekts

Ein Shoebox-Projekt ist eine Datei mit der Erweiterung **.prj**, in der die relevanten Einstellungen (engl. *settings*) vermerkt werden. Gehen Sie folgendermaßen vor:

- Erstellen Sie im Ordner *Shoebox* ein Unterverzeichnis (Ordner) namens *Esperanto*.
- Starten Sie das Shoebox-Programm.
- Legen Sie über Project, New in dem Esperanto-Ordner ein neues Projekt namens **esperanto.prj** an.

Einrichten einer neuen Datenbank

Für unser Esperanto-Projekt benötigen wir – wie für das Test- und das Latein-Projekt – ein Esperanto-Deutsches Wörterbuch, das wir als erstes anlegen wollen.

- Wählen Sie File, New und legen Sie in **Shoebox\Esperanto** eine Lexikodatei namens **esperanto.lex** an.
- Wählen Sie Speichern

Der Datenbank-Typ Es öffnet sich die Dialogbox mit der Beschriftung *Select Database Type*. Wählen Sie Add New Type.

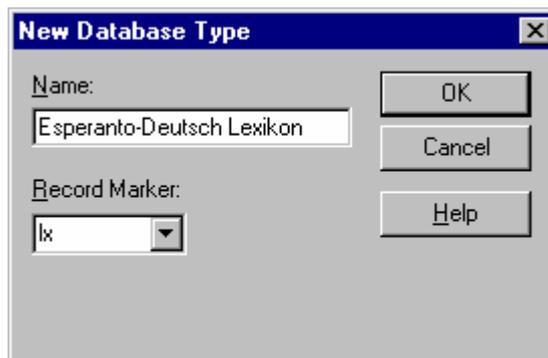
Es öffnet sich ein weiteres Dialogfenster mit dem Titel *Database Types*, welches alle verfügbaren Typen auflistet. In unserem Falle sollte die Liste leer sein.

Wählen Sie Add.

Es öffnet sich ein weiteres Dialogfenster mit der Bezeichnung *New Database Type*. Darin wird nach dem Namen für einen neuen Datenbank-Typ gefragt und nach der Datensatz-Markierung (*record marker*). Der Name muss nicht aus einem einzigen Wort bestehen.

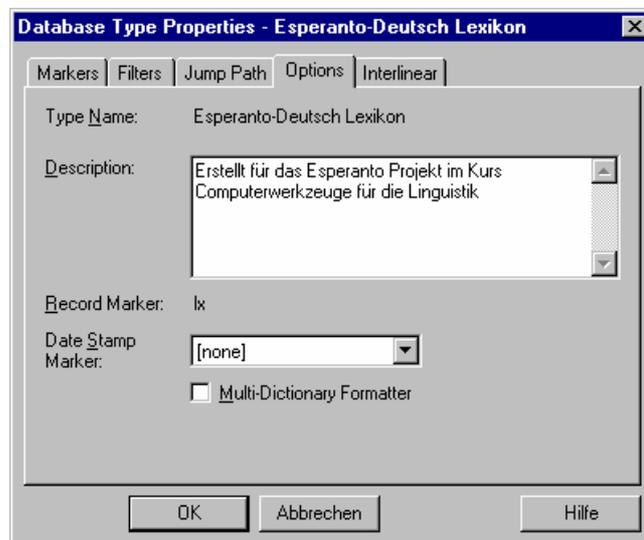
Geben Sie *Esperanto-Deutsch Lexikon* als Namen an.

Spezifizieren Sie *lx* (= Lexem) als Datensatz-Markierung (*Record Marker*).



Wählen Sie OK.

Es erscheint ein großes Dialogfenster mit Karteikarten und der Aufschrift *Database Type Properties* (Datenbank-Typ-Eigenschaften). Wir befinden uns in der Karteikarte *Options*. Das Programm fordert uns damit auf, eine Beschreibung (*Description*) des Datenbank-Typs zu erstellen.¹



Geben Sie den in der Graphik angezeigten Text als Beschreibung ein.

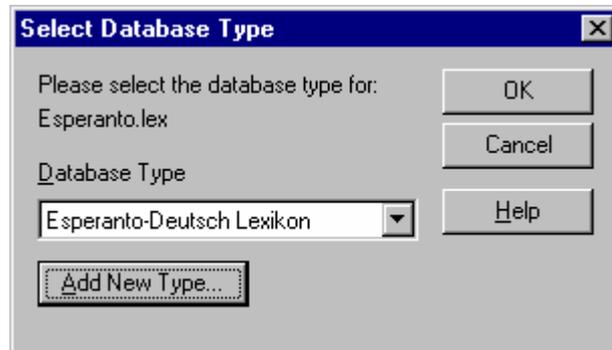
(Der *date stamp marker* kann jetzt noch nicht spezifiziert werden, weil bis jetzt nur die Datensatz-Markierung (*record marker*) definiert worden ist. Wir werden später hierher zurückkommen, nachdem einige weitere Markierungen definiert worden sind.)

¹ Diese Information kann nützlich sein, wenn man nach einer längeren Unterbrechung in seine Arbeit wieder aufnehmen will, oder wenn andere mit den Daten arbeiten wollen.

Die Bedeutung der anderen Karteikarten braucht uns an dieser Stelle noch nicht zu interessieren. Wir werden dies später nachholen.

- Schließen Sie den *Database Type Properties*-Dialog durch Anklicken von OK.
- Schließen Sie den *Database Type*-Dialog durch Anklicken von Close.

Wir befinden uns jetzt wieder in dem kleinen Dialogfenster *Select Database Type*.



- Wählen Sie OK.

Dies bringt uns zum Shoebox Hauptfenster zurück. Es enthält ein Fenster mit dem Titel *Esperanto.lex* und zeigt einen leeren Datensatz mit einer **!x** Markierung

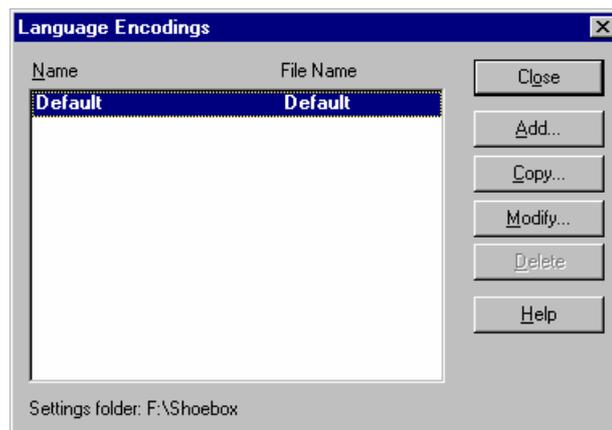
Die Sprach- Kodierung

*Name, Tastatur, und
Standard-Zeichensatz*

Das Esperanto-Alphabet enthält, wie bereits festgestellt, eine Reihe von Sonderzeichen. An dieser Stelle besteht die Möglichkeit, einen passenden Zeichensatz sowie andere Spracheigenschaften zu spezifizieren, die in Datensatzfeldern verwendet werden können, die sich auf Esperanto beziehen. Diese Sprachkodierungs-Information (engl. *Language encoding information*) wird getrennt von der Datenbank-Typ-Information verwaltet.

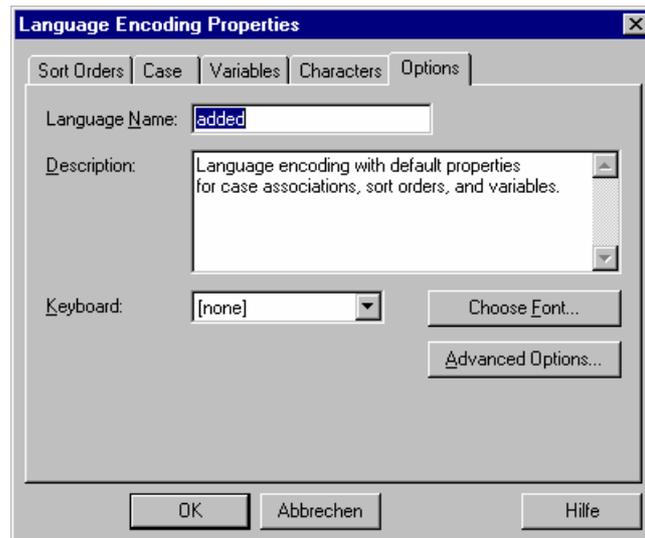
- Wählen Sie Project, Language Encodings.

Es öffnet sich ein Dialogfenster mit dem Titel *Language Encodings*, das eine Liste von verfügbaren Sprachkodierungen enthält. Die einzige zunächst verfügbare Spezifikation ist eine Standardbeschreibung (*Default*), die in den bisherigen Projekten auch verwendet worden ist.



- Wählen Sie Add.

Dies bringt uns zu einem "Karteikasten" mit dem Titel *Language Encoding Properties*, mit der Karteikarte *Options* als erstem Angebot.

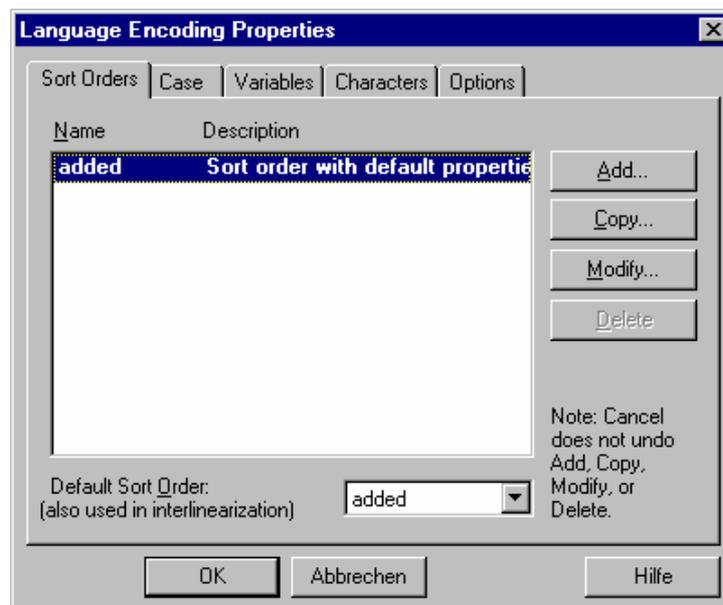


- Tragen Sie *Esperanto* als Sprachnamen (*Language Name*) ein.
- Tragen Sie den Text *Erstellt für das Esperanto-Projekt im Kurs Morphologie mit Shoebox* als Beschreibung (*Description*) ein.

An dieser Stelle würde es sich anbieten, einen Standardzeichensatz für die jeweilige Sprache zu spezifizieren. Dieser Zeichensatz würde dann in allen Feldern verwendet werden, die Daten dieser Sprache enthalten. Wir werden für unser Esperanto-Projekt zunächst von der Spezifikation eines solchen Zeichensatzes absehen.

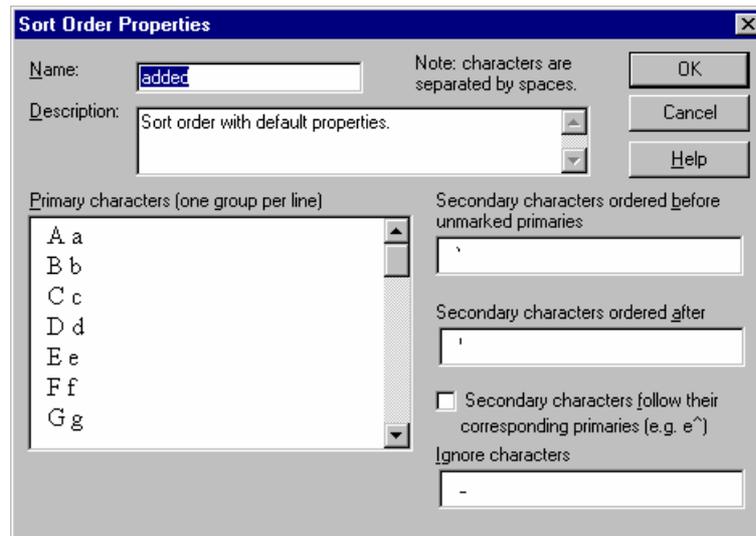
Die Sortierreihenfolge Hier geht es darum, für die alphabetische Auflistung besonders von Wörterbucheinträgen eine Reihenfolge festzulegen. Verschiedene Sprachen kennen dafür verschiedene Konventionen, natürlich abhängig vom jeweiligen Phoneminventar. Während im Deutschen das Alphabet mit *a, b, c* beginnt, haben wir im Russischen (phonetisch betrachtet) die Reihenfolge *a, b, v* (а, б, в).

- Wählen Sie die Kartei *Sort Orders*.



Wir stellen fest, dass das Programm für uns eine Standard-Sortierreihenfolge erzeugt hat und ihr den Namen *added* gegeben hat. Wir müssen hier natürlich Änderungen für Esperanto vornehmen, daher:

- Wählen Sie *Modify*.



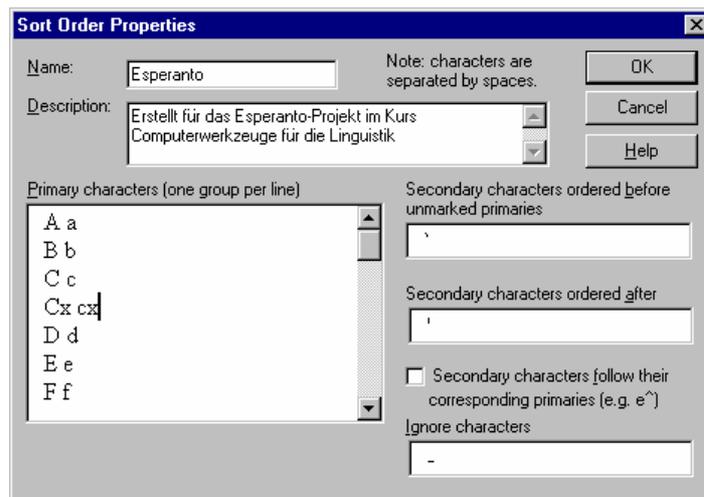
Sie sehen jetzt ein großes Dialogfenster namens *Sort Order Properties*.

- Ändern Sie den Namen in *Esperanto* um.
- Ändern Sie die Beschreibung (*Description*) in *Erstellt für das Esperanto-Projekt im Kurs Morphologie mit Shoebox*.

HINWEIS: Wenn Sie Schwierigkeiten mit der Terminologie haben, die hier verwendet wird, wählen Sie Help. Dort finden Sie Beispiele für die verschiedenen Zeichentypen (*primary*, *secondary*, etc.).

Das Programm bietet als Standard die Sortierreihenfolge des Englischen an. Diese gilt grundsätzlich auch für Esperanto. Wir haben aber gesehen, dass Esperanto zusätzlich Zeichen enthält (die wir mit Hilfe des Zusatzzeichens *x* kodieren). Wir müssen diese Zeichen daher hinzufügen, beispielsweise \hat{C} , \hat{c} in der Notierung *Cx*, *cx*.

- Plazieren Sie die Schreibmarke im Fenster *Primary characters* hinter das Paar *C*, *c* und drücken Sie die Eingabetaste (*Enter*).
- Es wird eine Neue Zeile eingefügt in der Sie dann das Zeichenpaar *Cx cx* eintragen können.



- Geben Sie auf die gleiche Weise alle übrigen Sonderzeichen von Esperanto ein.
- Entfernen Sie alle Zeichen, die **nicht** zum Esperanto-Alphabet gehören (z.B. *Q* oder *X*).

Zum Abschluss führen Sie bitte noch folgendes aus:

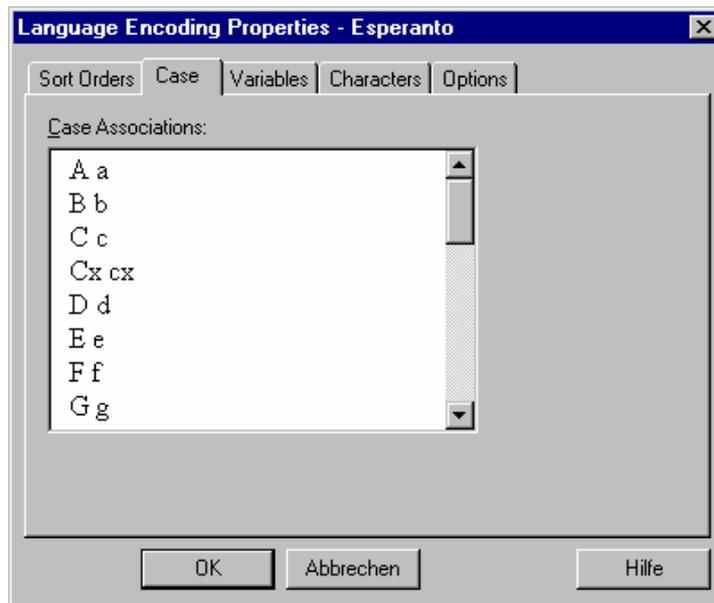
- Löschen Sie alle Zeichen aus *Secondary characters ordered before...*
- Löschen Sie alle Zeichen aus *Secondary characters ordered after...*

- Wählen Sie OK um zum *Language Encoding Properties* Dialogfenster zurückzukehren.

Unterhalb der Liste der verfügbaren Sortierreihenfolgen befindet sich ein Eingabefeld mit der Bezeichnung *Default Sort Order*. Hier können Sie die Standardreihenfolge einstellen, die für die fragliche Sprache gilt. Dies ist die Sortierreihenfolge, die für jedes Esperanto-Datensatzfeld verwendet wird, das beim Sortieren als sekundäres Sortierfeld gewählt wird. (Dazu später mehr). Es wird *Esperanto* angezeigt, was ja auch richtig ist.

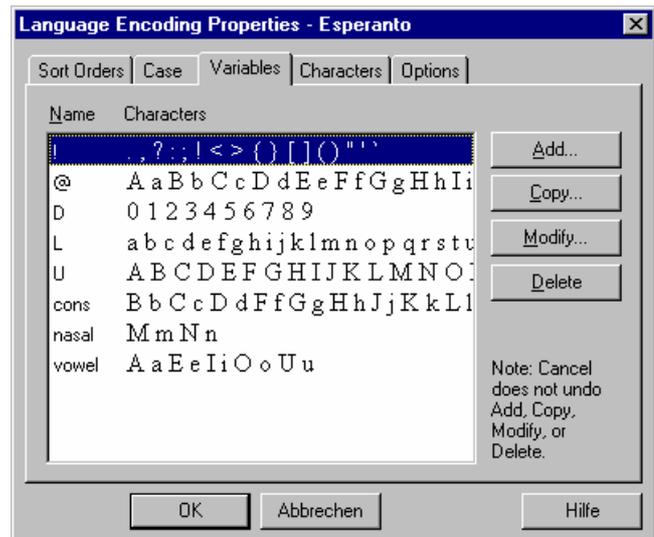
Die Case Karteikarte Hier geht es um die Spezifizierung der Groß-Klein-Schreibung in Sprachen, für die dies relevant ist. (Im Arabischen gibt es keine Groß-Klein-Schreibung).²

- Wählen Sie die Karteikarte Case.
- Geben Sie hier ebenfalls die Sonderzeichen des Esperanto ein.



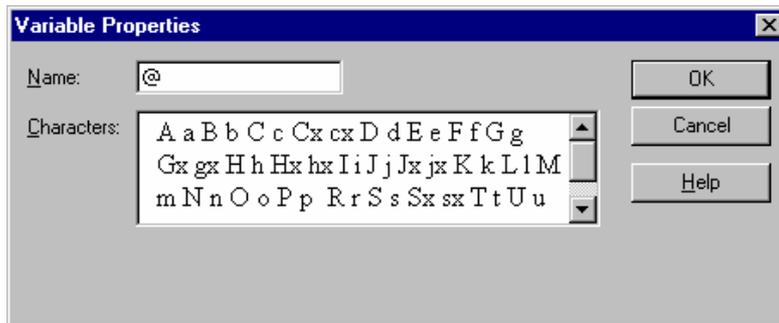
Alle Zeichen, für die es eine systematische Beziehung zwischen Groß- und Kleinbuchstaben gibt, müssen hier eingetragen werden.

Die Variables Kartei *Variablen* sind Namen für vom Nutzer definierte Zeichenklassen, die beim Durchsuchen der Datenbestände nützlich sein können. Eine mögliche Zeichenklasse für Esperanto wäre z.B. die Klasse der Vokale A a E e I i O o U u.



² Die Bezeichnung *case* kommt von *upper case* (Großbuchstaben) vs. *lower case* (Kleinbuchstaben); diese wiederum haben damit zu tun, dass früher die zum Setzen verwendeten Buchstaben in Kästen (*case*) einsortiert waren und zwar die Großbuchstaben in einem oberen Kasten und die Kleinbuchstaben in einem unteren Kasten.

- Wählen Sie die Karteikarte Variables.
- Wählen Sie die Variable @. Diese steht für einen beliebigen Buchstaben aus dem Zeichenvorrat. Wir müssen daher auch hier die Sonderzeichen ergänzen.



- Wählen Sie Modify.³
- Tragen Sie die Sonderzeichen des Esperanto entsprechend der Sortierreihenfolge in die Zeile ein und entfernen Sie alle nicht dazugehörigen Zeichen.
- Wählen Sie OK.

So definierte Variablen können in Ansichtsfiltren und im Suchbefehl (*Find command*) verwendet werden. Alle können modifiziert und neue können hinzugefügt werden. Da jede Variable eine spezifische Zeichenklasse definiert, sind sie Bestandteil der Sprachkodierung.

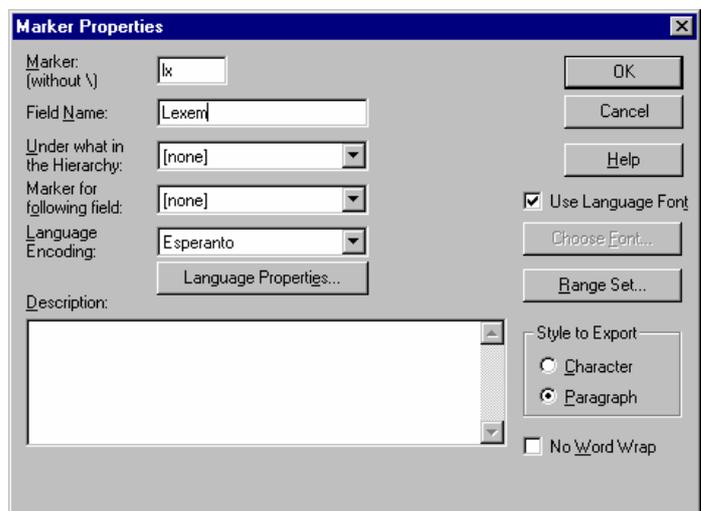
- Ändern Sie in analoger Weise die übrigen Variablen. Ändern Sie dabei gleichzeitig die Namen der Variablen, *cons* in *kons*, *vowel* in *vokal*, *D* in *Z* (Ziffern), *L* in *K* (Kleinbuchstaben), *U* in *G* (Großbuchstaben).
- Wählen Sie OK zur Bestätigung der Änderungen im Karteikasten *Language Encoding Properties*.
- Schließen Sie *Language Encodings* mit Close.

Damit sind die Vorbereitungen abgeschlossen, und wir können mit der eigentlichen Arbeit beginnen, dem Eintragen von Datensätzen in die Datenbank.

Eintragen eines Datensatzes

Alle Dialogfenster sollten inzwischen geschlossen sein. Es sollte nur noch das leere Datenfenster mit dem Titel *Esperanto.lex* sichtbar sein. Im Markierungsfenster sollte die Markierung **lx** zu sehen sein.

- Öffnen Sie durch Anklicken der Markierung **lx** mit der rechten Maustaste das Dialogfenster für die Markierungseigenschaften (*Marker Properties*) der Markierung **lx**.



³ Wählen Sie Help im Dialogfenster *Variable Properties*, um eine Liste der Standardvariablen und deren übliche Bedeutung zu erhalten. Die Variablennamen können geändert werden. Dieser Hilfeabschnitt beschreibt auch die vordefinierten Variablen (*Built-in variables*).

Einstellung der
Sprachkodierung
(language encoding) der
Markierung

- Geben Sie *Lexem* als Feldnamen an.
- Wählen Sie *Esperanto* aus der Sprachenliste (*Language Encoding*).
- Beenden Sie mit OK.

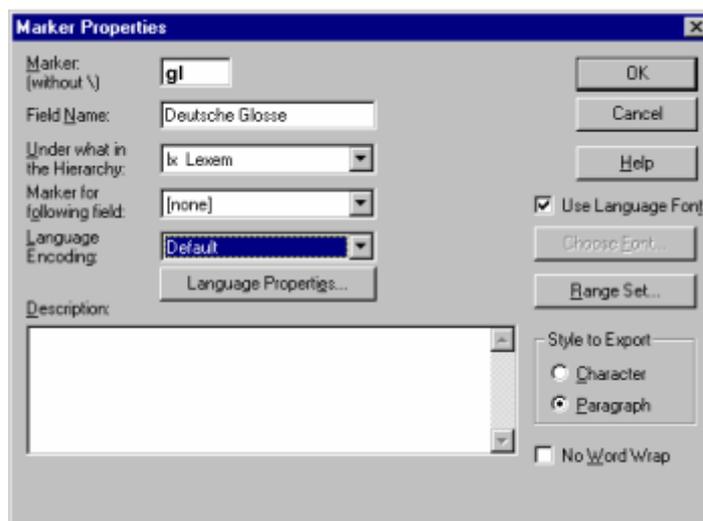
Die Schreibmarke steht jetzt wahrscheinlich im leeren Datenfeld und Sie können das erste Wort eingeben.

- Geben Sie *hundo* als ersten Lexikoneintrag ein.

Bis jetzt besteht unser Datensatz nur aus einem Feld. Damit kann man nicht viel anfangen. Wir würden nicht mehr als eine Liste von Esperantowörtern mit keinen weiteren Informationen erhalten.

- Wir fügen als erstes ein Feld ein, das es uns erlaubt eine deutsche Übersetzung anzugeben. Im Kontext der Interlinearisierung von Texten spricht man hier häufig von *Glossen*. Wir führen daher als Markierung die Abkürzung **gl** (für Glosse) ein.
- Drücken Sie die Eingabetaste (*Enter*) um Platz für den Eintrag zu schaffen und geben Sie dann **gl** gefolgt von einem Zwischenraum (Leertaste) ein.

Sie werden dann gefragt, ob Sie die neue Markierung der Liste hinzufügen wollen. Bestätigen Sie dies mit Ja.



- Wählen Sie Ja.

Damit wird erneut das Dialogfenster für die Markierungseigenschaften (*Marker Properties*) aufgerufen (mit **gl** als Markierung).

- Tragen Sie *Deutsche Glosse* als Feldnamen ein.
- Wählen Sie wie vorgeschlagen Default als Sprachkodierung (*Language Encoding*).

Eigentlich sollte man dem Datenfeld *Deutsche Glosse* eine Sprachkodierung für Deutsch zuweisen. Eine solche haben wir aber bisher nicht definiert. Für den Augenblick geben wir uns daher mit der Standarddefinition zufrieden.

- Wählen Sie OK.

Damit kehren wir zum Esperanto-Lexikon-Fenster zurück.

- Geben Sie *Hund* als Glosse für *hundo* ein.

Wir wollen noch weitere Feldnamen hinzufügen, bevor wir ein Datensatzmuster für diese Datenbank erstellen. Definieren Sie nach dem oben beschriebenen Verfahren weitere Datenfelder:

- Markierung: **\ps** Feldname: *Part of Speech* Sprache: *Default*
- Markierung: **\dt** Feldname: *Datum* Sprache: *Default*

Wir machen jetzt das Feld **\dt** zu einem Datumsstempel (*date stamp*), was bedeutet, dass bei einem neuen Eintrag automatisch das aktuelle Datum eingetragen wird.

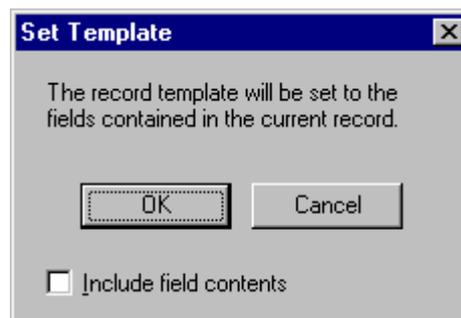
- Wählen Sie Database, Properties (beachten Sie, dass in der Liste jetzt alle neuen Feldmarkierungen aufgeführt sind).
- Wählen Sie die Karteikarte Options.
- Aktivieren Sie die Liste Date Stamp Marker und wählen Sie **dt Datum** aus der Liste.
- Bestätigen Sie mit OK.

Damit haben wir den Grundstock für die ersten Schritte eines Esperanto-Lexikons. Wir werden später noch eine Reihe von Eigenschaften hinzufügen müssen, um ernsthaft damit arbeiten zu können.

Erstellen eines neuen Datensatzmusters

Als nächstes wollen wir ein Muster für einen Datensatz in unserem Esperanto-Deutsch-Lexikon erstellen. Es handelt sich dabei um eine Zusammenstellung der Markierungen, die automatisch eingefügt werden sollen, wann immer ein neuer Datensatz erstellt werden soll.

- Wählen Sie Database, Template.



- Wählen Sie OK und das Dialogfenster verschwindet.

Einfügen des zweiten Datensatzes

Führen Sie folgende Schritte aus, um den zweiten Datensatz zu generieren,

- Wählen Sie Database, Insert Record oder benutzen Sie das Tastaturkürzel (Strg+N). Es öffnet sich ein Dialogfenster mit dem Titel *Insert Record*.
- Geben Sie *patro* als einzufügender Datensatz an.
- Wählen Sie OK.

Das Programm kehrt dann zum Lexikonfenster zurück, das den Lexemeintrag *patro* bereits enthält. Die Schreibmarke befindet sich auf der **\gl**-Linie. Beachten Sie auch, dass das Muster die Markierungen **\gl**, **\nt**, und **\dt** automatisch eingefügt hat.



- Geben Sie *Vater* als Glosse ein.

Der Datumsstempel ist noch nicht ausgefüllt. Dies geschieht erst wenn den Datensatz verlassen hat.

- Klicken Sie auf den Previous Record Schalter um zum Eintrag *hundo* zu gelangen (hier ist der Datumsstempel ebenfalls leer, weil die Markierung erst nach dem Einfügen des ersten Datensatzes definiert worden ist.
- Fügen Sie N im Feld \ps ein.
- Klicken Sie dann auf den Next Record Schalter um zu *patro* zurückzukehren. Das Datumsfeld sollte jetzt mit dem heutigen Datum ausgefüllt sein.
- Klicken Sie erneut auf Previous Record. Jetzt sollte auch hier das Datumsfeld einen Eintrag haben.

Gegeben seien folgende Sätze mit ihren Übersetzungen. Tragen Sie die jeweiligen Wörter in die Datenbank ein.

La patro estas en la domo. Der Vater ist im Haus.

En la salono Sinjoro Lang legas kaj fumas. Im Salon liest und raucht Herr Lang.

Ekster la domo estas hundejo por la hundo. Außerhalb des Hauses ist die Hundehütte für den Hund.