



Programm

- Was ist Shoebox?
 - ◇ Allgemeines zu Datenbankprogrammen
 - ◇ Shoebox-Datenbanken
- Wieso ist Shoebox für Linguisten interessant?
 - ◇ Lexikondatenbanken exportieren
 - ◇ Interlinearisierung
- Schnuppershoebox
 - ◇ Lexikon
 - ◇ Interlinearisierung
 - ◇ Interaktion zwischen Text & Lexikon
- Shoebox im Studium



Was ist Shoebox?

Shoebox ist im Kern ein **DATENBANKPROGRAMM**.

Datenbankprogramme dienen im wesentlichen dazu, Daten bzw. Information digital...

- ❖ strukturiert zu erfassen und zu verwalten um so schnellen Zugriff auf sie zu ermöglichen
- ❖ anderen Anwendungsprogrammen zur Verfügung zu stellen
- ❖ zu neuen Daten weiterzuverarbeiten.

Um was für Daten bzw. Information es sich im einzelnen handelt, ist individuell ganz unterschiedlich und hängt jeweils von den **Aufgabenstellungen** ab, die die Nutzer eines Datenbankprogrammes am Wickel haben.

Datenbanken sind aus der Industriegesellschaft nicht mehr wegzudenken - sei es die Verwaltung der Gehälter in öffentlichen Einrichtungen, die Organisation der Finanzbuchhaltung in Konzernen, die Erfassung von Mitgliedern und Angehörigen in Schulen, Unis, Bibliotheken, die Buchführung in einem Supermarkt usw. usf.: immer wenn es um die Verarbeitung größerer Mengen von Daten geht, sind entsprechende Programme im Einsatz. Diese sind i.d.R. für den bestimmten Anwendungsbereich maßgeschneidert.

Auch für 'individuelle' oder 'private' User stehen Programme zur Verfügung (Access, DBase, Filemaker), um Daten und Information zu organisieren und zu verwalten (Literaturdatenbank, Adressdatenbank usw.)



Was ist Shoebox?

Die Information innerhalb einer Datenbank ist nach spezifischen **Kriterien** inhaltlich sortiert. Das bedeutet, dass in einer Menge von Daten ganz bestimmte **Kategorien** definiert werden, denen dann jeweils eine Gruppe von Daten zugeordnet ist.

Beispiel 'Personendatenbank'

Vorname	Name	Fach	Uni
John	Bateman	Anglistik	Bremen
Winfried	Boeder	Anglistik	Oldenburg
Thomas	Stolz	Linguistik	Bremen
Karl Heinz	Wagner	Linguistik	Bremen
Wolfgang	Wildgen	Germanistik	Bremen



Was ist Shoebox?

Die Information innerhalb einer Datenbank ist nach spezifischen **Kriterien** inhaltlich sortiert. Das bedeutet, dass in einer Menge von Daten ganz bestimmte **Kategorien** definiert werden, denen dann jeweils eine Gruppe von Daten zugeordnet ist.

Beispiel 'Mini-Lexikon'

Morphem	Kategorie	Glosse
boy	Noun	junge
kick	Verb	tret-
the	Det	Def
-ed	Suffix	Past
-er	Suffix	Agnt



Was ist Shoebox?

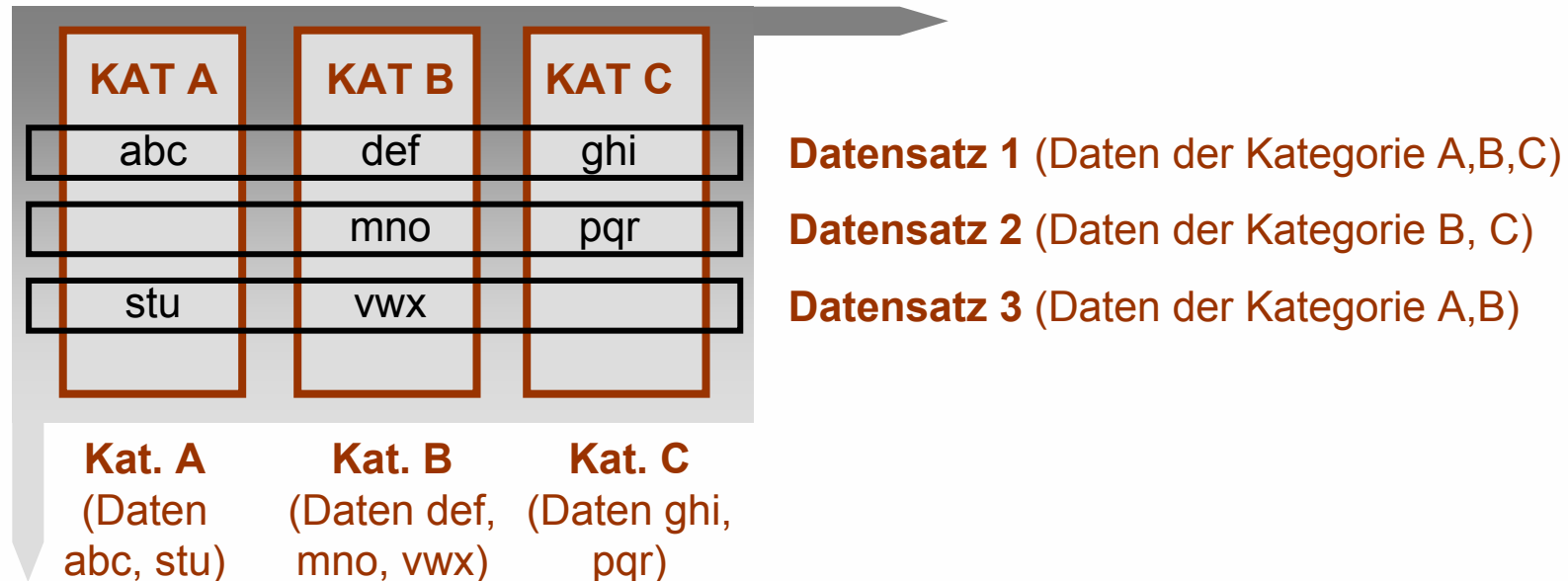
Die Information innerhalb einer Datenbank ist nach spezifischen **Kriterien** inhaltlich sortiert. Das bedeutet, dass in einer Menge von Daten ganz bestimmte **Kategorien** definiert werden, denen dann jeweils eine Gruppe von Daten zugeordnet ist.

Beispiel 'Bibliographie'

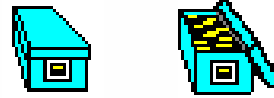
Autor	Titel	Jahr	Ort	Verlag
Dik, Simon	Functional Grammar	1978	Dordrecht	Foris
Lyons, John	Semantics	1977	Cambridge	CUP
Carnie, Andrew	Syntax	2002	Oxford	Blackwell
Anderson, John M.	On Case Grammar	1977	London	Croom Helm



Was ist Shoebox?



Wie die Beispiele zeigen, ist die Information in einer Datenbank auf zweierlei Art organisiert: zum einen sind Daten der gleichen Art einer **gemeinsamen Kategorie** zugeordnet; zum anderen bilden Daten unterschiedlichen Typs einen gemeinsamen **Datensatz (engl. Record)**. Jeder einzelne Datensatz unterteilt sich also in Informationseinheiten unterschiedlicher Kategorie. Anders ausgedrückt: jeder Datensatz in einer Datenbank besteht aus einer Menge von **DATENFELDERN**, die sich ihrerseits aus der Angabe über die Kategorie und dem konkreten Inhalt zusammensetzen.



Was ist Shoebox?

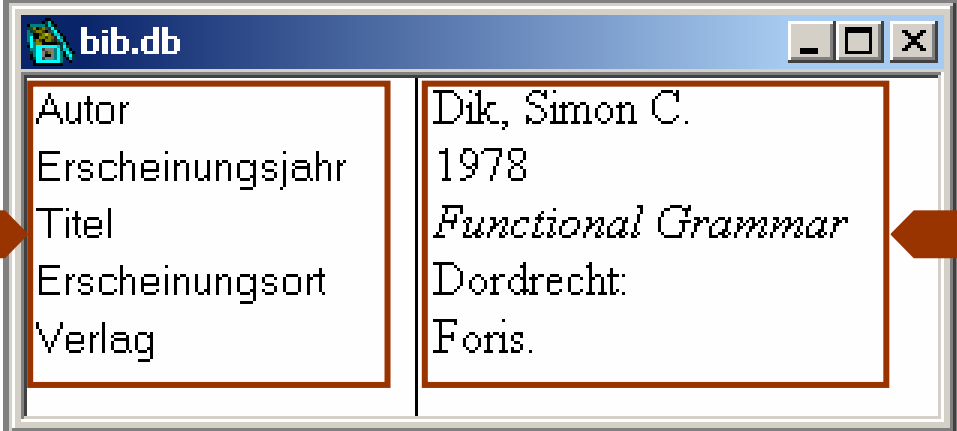
Datenbanken in aller Kürze:

Eine Datenbank ist eine Menge von **Datensätzen** (Records), die ihrerseits aus einer Menge von Datenfeldern bestehen.

Ein **Datenfeld** ist die Zuordnung einer **konkreten Information** zu einer Kategorie, die als sogenannter **Feldname** angegeben ist.

Die Anzahl der Datenfelder und deren Bezeichnung, sprich die Feldnamen, werden vom User festgelegt und sind i.d.R. invariant. Die in einem Datenfeld kodierte konkrete Information variiert in jedem Datensatz.

Beispiel: Datensatz einer Bibliographie-Datenbank

<p>Invariant - jeder Eintrag setzt sich aus den Datenfeldern Autor - Verlag zusammen</p>	 <table border="1"><tr><td>Autor</td><td>Dik, Simon C.</td></tr><tr><td>Erscheinungsjahr</td><td>1978</td></tr><tr><td>Titel</td><td><i>Functional Grammar</i></td></tr><tr><td>Erscheinungsort</td><td>Dordrecht:</td></tr><tr><td>Verlag</td><td>Foris.</td></tr></table>	Autor	Dik, Simon C.	Erscheinungsjahr	1978	Titel	<i>Functional Grammar</i>	Erscheinungsort	Dordrecht:	Verlag	Foris.	<p>Individuell - die den Feldern jeweils zugeordneten Daten variieren mit jedem Eintrag</p>
Autor	Dik, Simon C.											
Erscheinungsjahr	1978											
Titel	<i>Functional Grammar</i>											
Erscheinungsort	Dordrecht:											
Verlag	Foris.											



Shoebox-Datenbanken

Jede Shoebox-Datenbank hat spezifische Eigenschaften. Dazu gehören ua. Anzahl und Kategorie der Datenfelder.

Die Kategorie des Datenfeldes trägt einen entsprechenden Feldnamen - im Bibliographie-Beispiel wie gesehen ua. 'Autor', 'Titel' usw. In der Anwendung, d.h. bei der Erstellung der Datenbank und deren Bearbeitung, wird allerdings nicht mit dem vollen Namen des Feldes gearbeitet, sondern stattdessen mit einem Kürzel, dem sogenannten **MARKER** oder **FIELD-MARKER**. Dieses Konzept spielt in der Arbeit mit Shoebox eine wichtige Rolle.

bib.db	
Autor	Dik, Simon C.
Erscheinungsjahr	1978
Titel	<i>Functional Grammar</i>
Erscheinungsort	Dordrecht:
Verlag	Foris.

Datensatz mit Feldnamen

bib.db	
\at	Dik, Simon C.
\ej	1978
\tt	<i>Functional Grammar</i>
\eo	Dordrecht:
\vl	Foris.

Datensatz mit Markern

Autor	\at
Erscheinungsjahr	\ej
Titel	\tt
Erscheinungsort	\eo
Verlag	\vl

Feldname Marker

Bei der Erstellung einer neuen Datenbank obliegt die Spezifizierung der darin vorhandenen Marker sowie die Zuordnung Marker/Feldname dem Anwender. Es ist allerdings anzumerken, dass es bestimmte Konventionen bezüglich der Markerbezeichnungen gibt.



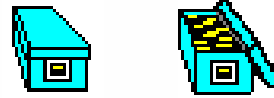
Shoobox-Datenbanken

Die Eigenschaften einer Shoebox-Datenbank - also z.B. Angaben darüber, wieviele Felder es potentiell geben kann, durch welche Marker diese Felder identifiziert werden, welche Feldnamen den Markern zugeordnet sind usw. - werden in Form eines **DATENBANKTYP** notiert. Dieser ist in einer eigenen Datei abgespeichert.

Bei der Erstellung einer neuen Datenbank muss für diese entweder ein neuer Datenbanktyp angelegt oder ihr ein bereits existierender Datenbanktyp zugewiesen werden: **Jede Shoebox-Datenbank rekurriert auf einen Datenbanktypen.**

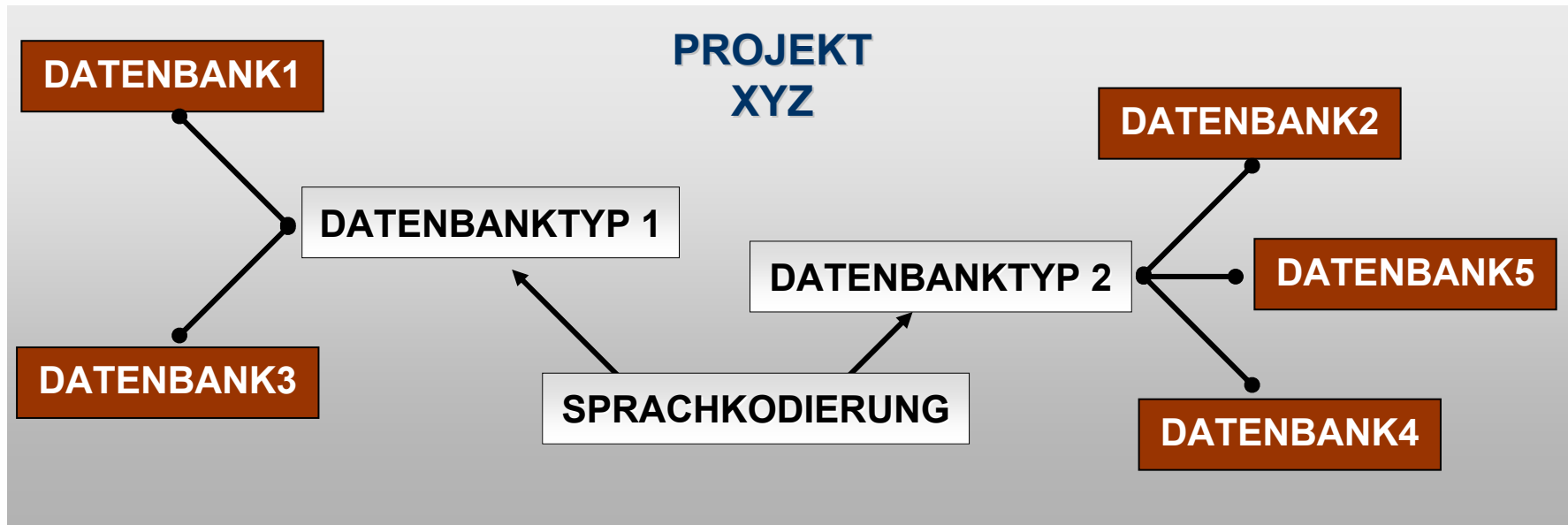


Dabei ist nicht ausgeschlossen, dass verschiedene Datenbanken auf demselben Datenbanktyp basieren.



Shoebox-Projekte

Inhaltlich zusammengehörige Datenbanken, die diesen zugrundeliegenden Datenbanktypen, ggf. Sprachkodierungen etc. werden in Shoebox in sogenannten **Projekten** organisiert und verwaltet.



Datenbanken, Datenbanktypen und Co sind **IMMER** in Projekten organisiert - **selbst dann, wenn nur eine einzige Datenbank erstellt werden soll**, muss für diese ein Projekt angelegt werden.



Shoebox-Projekte

Die mit einem jeden Projekt abgespeicherte Information umfasst nicht nur die Angaben in neu angelegten bzw. modifizierte Dateien (Datenbanken, Datenbanktypen, Sprachkodierungen) usw.

In einem Projekt finden sich neben den von den Usern produzierte Dateien - also den Datenbanken - auch weitere, von Shoebox automatisch angelegte Dateien. Der einem bzw. mehreren Datenbanken zugrundeliegende Datenbanktyp wird automatisch in einer eigenen Datei abgespeichert; ebenso die Sprachkodierungen.

Shoebox 'merkt' sich beim Verlassen des Programmes darüberhinaus, welches Projekt zuletzt bearbeitet wurde, welche Datenbank(en) des Projekts geöffnet war(en) und welche Fenster in welcher Größe und Lage auf dem Monitor angezeigt wurden. Anders ausgedrückt: der User findet seine Arbeitsumgebung in genau der Form vor, in der er sie verlassen hat.

Anmerkung: dieser Service erweist sich als problematisch, wenn Projekte kopiert oder verschoben werden.



Wieso ist Shoebox für Linguisten interessant?

Shoebox ist insofern mehr als ein 'reines' Datenbankprogramm, als es für einige der spezifischen Bedürfnisse von Linguisten zugeschnitten ist.

Hier sind insbesondere zwei Features relevant, nämlich

- a) der **Export einer Lexikondatenbank in ein anderes Format**
- b) die **Interlinearisierung von Texten** mit Bezug auf eine Shoebox-Lexikondatenbank

Zur Erinnerung:

Datenbankprogramme dienen im wesentlichen dazu, Daten bzw. Information digital...

- a) strukturiert zu erfassen und zu verwalten um so schnellen Zugriff auf sie zu ermöglichen
- b) **anderen Anwendungsprogrammen zur Verfügung zu stellen**
- c) zu neuen Daten weiterzuverarbeiten.



Wieso ist Shoebox für Linguisten interessant?

LEXIKONDATENBANKEN EXPORTIEREN

Im Grunde sind Shoebox-Datenbanken nicht mehr als einfache Textdateien. Über spezielle Zusatzprogramme aber können die Daten einer Shoebox-Datenbank **individuell formatiert** in ein anderes Textformat wie z.B. .rtf oder .doc konvertiert werden.

In diesem Kontext kann Shoebox also als ein Programm für die Erstellung von Lexika gesehen werden.

Shoebox-Datensatz

\lx balun
\ps n
\ge side ; remainder ; some
\lf Idiom = ai balun
\le casket (lit. 'its wooden sides')
\mr balu-n
\cf balu
\ce part



rtf oder doc-Exzerpt

balun *n.* side, remainder, some. *Idiom:* **ai balun** 'casket (lit. 'its wooden sides')'. *Morph:* **balu-n**. *See:* **balu** 'part'.

[polymorphemic lexeme; identifying **balu** and **-n**]



Wieso ist Shoebox für Linguisten interessant?

Shoebox ist insofern mehr als ein 'reines' Datenbankprogramm, als es für einige der spezifischen Bedürfnisse von Linguisten zugeschnitten ist.

Hier sind insbesondere zwei Features relevant, nämlich

- a) der **Export einer Lexikondatenbank in ein anderes Format**
- b) die **Interlinearisierung von Texten** mit Bezug auf eine Shoebox-Lexikondatenbank

Zur Erinnerung:

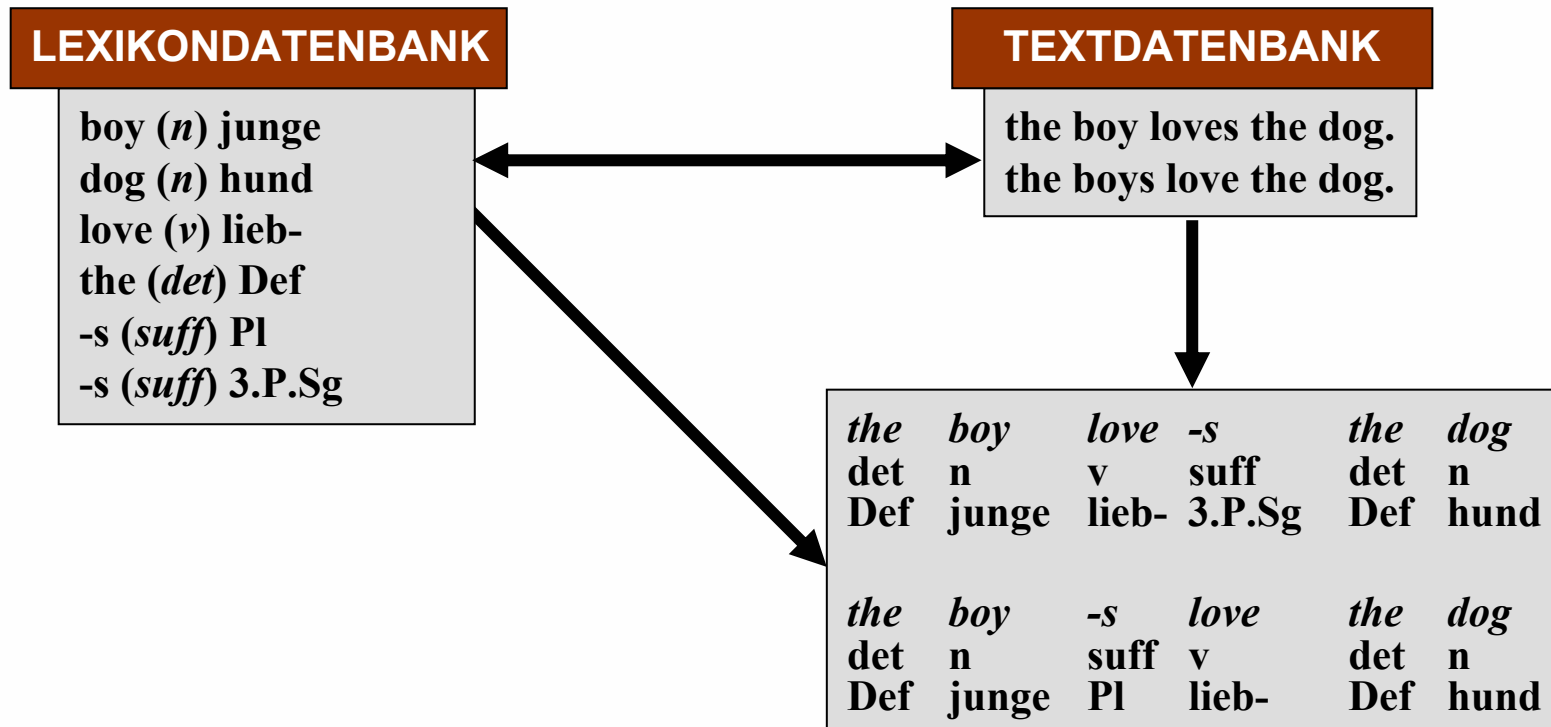
Datenbankprogramme dienen im wesentlichen dazu, Daten bzw. Information digital...

- a) strukturiert zu erfassen und zu verwalten um so schnellen Zugriff auf sie zu ermöglichen
- b) anderen Anwendungsprogrammen zur Verfügung zu stellen
- c) **zu neuen Daten weiterzuverarbeiten.**



Wieso ist Shoebox für Linguisten interessant?

Besonders interessant für Linguisten ist die Möglichkeit, einen Text auf der Basis einer Lexikondatenbank morphologisch zu parsen und automatisch eine **INTERLINEARVERSION** dieses Textes erzeugen zu lassen





In aller Kürze ...

Shoebox ist ein Programm für die Erfassung, Verwaltung und Bearbeitung von **Daten**.

Die Daten sind in **Datenbanken** organisiert.

Datenbanken sind in einzelne **Datensätze (Records)** unterteilt, die sich aus einer Menge von **Datenfeldern** konstituieren.

Ein Datenfeld besteht aus einer **Feldkennzeichnung (Marker)**, die durch Backslash + Buchstabenkürzel charakterisiert ist, und den in den einzelnen Datensätzen eingegebenen Daten. Jedem Marker kann ein (verständlicher) **Feldname** zugewiesen werden.

Jede Shoebox-Datenbank basiert auf einem **Datenbanktyp**, in dem Aufbau und Struktur der Datenbank determiniert ist. Verschiedenen Datenbanken kann derselbe Datenbanktyp zugewiesen sein.

Zusammengehörige Datenbanken, Datenbanktypen etc. werden in Shoebox in einem **Projekt** erfasst und verwaltet.

Für Linguisten ist Shoebox ua. deshalb interessant, weil es **Lexikonproduktion** und **Interlinearisierung** von Texten unterstützt.



Audience Participation I: Ein Mini-Lexikon erweitern

Melden Sie sich wie folgt am Rechner an:

Benutzername: pckurs

Kennwort: pckurs

Starten Sie, wenn der Rechner hochgefahren ist, das Programm SHOEBOX.

(START > LINGUISTIK-WERKZEUGE > SHOEBOX 4.0 > SHOEBOX 4.0 FOR WIN 95)

Shoebox öffnet sich mit dem Projekt test.prj. Diese Information befindet sich am unteren rechten Rand des Shoebox-Fensters.

Auf dem Monitor ist die Datenbank **lexikon1.lex** als Fenster mit dem Datensatz für das Lexem *boy* zu sehen.

- A) Finden Sie heraus, welche Datensätze in dieser Datenbank enthalten sind
(**Button 'Browse View', Pfeiltasten**)
- B) Finden Sie heraus, wie die Feldnamen der Marker lauten
(**Menü 'View', Befehl 'Field Names'**)
- C) Ergänzen Sie die Lexikondatenbank durch Einträge für die Items
dog, bark, with, hat, blue, in, garden, sleep
(**Menü 'Database', Befehl 'Insert Record' oder Tastenkombination STRG+N**)



Audience Participation II: Eine Mini-Interlinearisierung

Öffnen Sie über das Menü **File > Open** die Textdatenbank **text1.txt**. Stellen Sie die Fenstergröße dieser Datei so ein, dass diese neben der Lexikodatei zu sehen ist.

- A) Finden Sie heraus, welche Datensätze in dieser Datenbank enthalten sind
- B) Finden Sie heraus, wie die Feldnamen der Marker lauten
- C) Gehen Sie in den ersten Datensatz (Reference 001), begeben Sie sich mit der Schreibmarke an den Anfang der Textzeile und klicken Sie auf den Interlinearisierungsbutton. Geben Sie anschließend in der Zeile 'Free Translation' (\f) eine entsprechende freie deutsche Übersetzung ein
- D) Interlinearisieren Sie auf dieselbe Art die Datensätze 002-006.



Audience Participation III: Lexikon-und-Text-Interaktion

Schließen Sie das Projekt test1.prj. (**Menü Project, Befehl Close**).

Sie werden gefragt, ob Sie ein neues Projekt anlegen, ein existierendes Projekt öffnen oder Shoebox verlassen wollen.

Wählen Sie die Option '**Open an existing project**'. Gehen Sie in das Verzeichnis **test2** und öffnen Sie die Datei **test2.prj**.

Auf dem Monitor sind die Fenster **lexikon2.lex** und **text2.txt** geöffnet.

- A) Machen Sie sich mit den Daten aus beiden Datenbanken vertraut
- B) Interlinearisieren Sie den ersten Datensatz der Textdatenbank und fügen Sie eine freie Übersetzung hinzu.
- C) Interlinearisieren Sie den zweiten Datensatz. Was fällt Ihnen auf?
Geben Sie Shoebox die gewünschte Information und fügen eine freie Übersetzung hinzu. Interlinearisieren Sie den dritten Datensatz.
- D) Interlinearisieren Sie den vierten Datensatz. Was fällt Ihnen auf?
Ergänzen Sie das Lexikon um die benötigte Information und interlinearisieren Sie diesen Datensatz fertig. Verfahren Sie bei den Datensätzen 005 und 006 ebenfalls nach diesem Verfahren.
- E) Interlinearisieren Sie den siebten Datensatz. Was fällt Ihnen auf?



Morphophonemische Prozesse in Shoebox

Auch 'kniffligere' Fälle beim morphologischen Parsen können mit Shoebox bearbeitet werden.

'Morphologische Alternanten'

*The boy is **insecure**.*

*The boy is **irresponsible**.*

*The boy is **illiterate**.*

*The boy is **impotent**.*

Negation *im, ir, il, in*:

im ⇒ b, p

il ⇒ l

ir ⇒ r

in ⇒ sonst



Morphophonemische Prozesse in Shoebox

Auch 'kniffligere' Fälle beim morphologischen Parsen können mit Shoebox bearbeitet werden.

'Zugrundeliegende Form'

The boy **kicked** the dog

The boy **went** home

The boy **found** the book

The boy **threw** the ball

Regelmäßiges Verb **kick**:

kicked ⇒ kick -ed

Unregelmäßige Verben **go, find, throw**:

went ⇒ go -ed

found ⇒ find -ed

threw ⇒ throw -ed



Shoebox im Studium

Shoebox ist ein nützliches Tool für Linguisten, wenn es um die Erfassung und Verarbeitung von (im wesentlichen morphologischen) Daten geht und kann also solches gewinnbringend im Studium eingesetzt werden, und zwar

als 'Werkzeug' wenn es - z.B. in Haus- oder Abschlußarbeiten - um die Verarbeitung größerer Datenmengen geht

als 'Mittel zur Erkenntnisgewinnung' wenn es darum geht, einen Gegenstandsbereich / eine Aufgabe wirklich gründlich und präzise und durchdacht zu beschreiben. Genau dazu ist man durch die Verwendung von Programmen wie Shoebox GEZWUNGEN

Im Unterschied zu Programmen wie Word, Powerpoint usw, die **SO TUN**, als ob man sie ohne jegliche Vorkenntnisse und Einweisungen benutzen kann (wobei dann häufig allergößter Mist entsteht), gibt es bei Shoebox keine derartige pseudo-Benutzerfreundlichkeit.

Mit anderen Worten: Shoebox ist nur unter größten Mühen selbst zu erlernen, es ist also anzuraten, sich einer entsprechende Unterweisung zu unterziehen.



Shoebox im Studium

Näheres über Shoebox (und andere interessante Programme) erfahren Sie in der LV

COMPUTERWERKZEUGE FÜR DIE LINGUISTIK, SOSE 2003, MI 15-17h, CIP-LABOR

deren Besuch dringend angeraten wird.