

Was ist Shoebox?

- ◇ Shoebox als Datenbankprogramm
- ◇ Wieso ist Shoebox für Linguisten interessant
- ◇ Schnupper-Shoebox
- ◇ Shoebox im Studium

Morphologie mit Shoebox

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

Morphologie mit Shoebox

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 zu bewerkstelligen haben.

Morphologie mit Shoebox

Was ist eine Datenbank?

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.: immer wenn es um die Verarbeitung größerer Mengen von Daten geht, sind Datenbank-Programme im Einsatz. Diese sind i.d.R. für den bestimmten Anwendungsbereich maßgeschneidert.

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

Morphologie mit Shoebox

Datenbankorganisation

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

Morphologie mit Shoebox

Datenbankorganisation

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

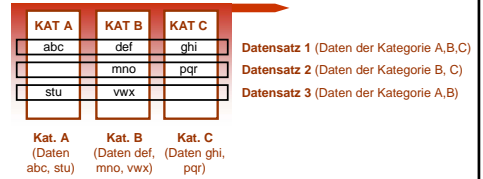
Was ist Shoebox?

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

Datenbankorganisation



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.

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 Anwender festgelegt und sind i.d.R. invariant. Die in einem Datenfeld kodierte konkrete Information variiert in jedem Datensatz.

Beispiel: Datensatz einer Bibliographie-Datenbank



Shoebox-Datenbanken

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

Die Kategorie des Datenfeldes trägt einen entsprechende 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.



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.

Shoebox-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**s notiert. Dieser ist in einer eigenen Datei abgespeichert.

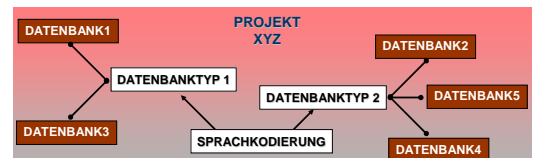
Bei der Erstellung einer neuen Datenbank muss für diese entweder ein neuer Datenbanktyp angelegt oder ihr ein bereits existierender Datenbanktyp zugewiesen werden: **Jede Shoebox-Datenbank rekuriert auf einen Datenbanktypen.**



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.



Morphologie mit Shoebox



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 Anwendern produzierten 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 Anwender findet seine Arbeitsumgebung in genau der Form vor, in der er sie verlassen hat.



Morphologie mit Shoebox



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

1. der **Export einer Lexikondatenbank in ein anderes Format**
2. die **Interlinearisierung von Texten** mit Bezug auf eine Shoebox-Lexikondatenbank

Zur Erinnerung:

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

1. strukturiert zu erfassen und zu verwalten um so schnellen Zugriff auf sie zu ermöglichen
2. **anderen Anwendungsprogrammen zur Verfügung zu stellen**
3. zu neuen Daten weiterzuverarbeiten.



Morphologie mit Shoebox



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
\if 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'.
 [polymorphic lexeme: identi-
 fying balu and -n]



Morphologie mit Shoebox



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

1. der **Export einer Lexikondatenbank in ein anderes Format**
2. die **Interlinearisierung von Texten** mit Bezug auf eine Shoebox-Lexikondatenbank

Zur Erinnerung:

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

1. strukturiert zu erfassen und zu verwalten um so schnellen Zugriff auf sie zu ermöglichen
2. **anderen Anwendungsprogrammen zur Verfügung zu stellen**
3. **zu neuen Daten weiterzuverarbeiten.**



Morphologie mit Shoebox



Interlinearisierung

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

LEXIKONDATENBANK

```

boy (n) junge
dog (n) hund
love (v) lieb-
the (det) Def
-s (suff) Pl
-s (suff) 3.P.Sg

```

TEXTDATENBANK

```

the boy loves the dog.
the boys love the dog.

```

```

the boy love -s the dog
det n v suff det n
Def junge lieb- 3.P.Sg Def hund

the boy -s love the dog
det n suff v det n
Def junge Pl lieb- Def hund

```



Morphologie mit Shoebox



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** konsituieren.
- ◇ 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.