

Der lexikografische Prozess von *elexiko*

Annette Klosa (Institut für Deutsche Sprache, Mannheim)

Grundsätzliches

- *elexiko* ist ein **Ausbauwörterbuch** (vgl. Lemberg 2001: 81), d. h., es ist (noch) nicht vollständig.
- *elexiko* ist direkt für die **Online-Publikation** geplant und realisiert.
- Der lexikografische Prozess für *elexiko* ist ein **computerlexikografischer Prozess** (vgl. Wiegand 1999: 233ff.), d. h., Computer werden durchgehend für die Erarbeitung und Publikation des Wörterbuches eingesetzt.

Grundsätzliches

- Folgende **Phasen des Prozesses** werden unterschieden:
 - Vorbereitung
 - Datensammlung
 - Computerisierung
 - Datenverarbeitung
 - Datenanalyse
 - Vorbereitung für Online-Release

Vorbereitungsphase

- **Wörterbuchskizze** und **Pilotstudien** zu den lexikografischen Angaben und der Stichwortliste sind im Band „Grundfragen der elektronischen Lexikographie. elexiko – Das online-Informationssystem zum deutschen Wortschatz“ (hrsg. von Ulrike Haß, Berlin/New York 2005) publiziert.
- Parallel sind **Musterartikel** erarbeitet worden.

Phase der Datensammlung

- Das **elexiko-Korpus** wurde zu Beginn der Arbeiten als virtuelles Korpus aus dem Deutschen Referenzkorpus des IDS konzipiert und zusammengestellt.
- **Sekundär- und Tertiärquellen**, d.h. andere Wörterbücher, Grammatiken etc., wurden parallel gesammelt.
- Mit der Beschaffung anderer Daten wie **Illustrationen** und **Tondateien** wurde erst später, während der Phase der Datenanalyse, begonnen.

Phase der Computerisierung

- *elexiko* konnte auf COSMAS II als **Korpusrecherche- und -analyswerkzeug** zurückgreifen, außerdem auf die Kookkurrenzdatenbank CCDB (vgl. Belica 2011ff.)
- Als **Datenbank** wird ORACLE eingesetzt, das schon am IDS für Grammis verwendet wurde.
- Als **XML-Editor** wurde zunächst XMetaL eingesetzt, inzwischen Oxygen.

Phase der Computerisierung

- Die Schnittstelle zwischen diesem und der Datenbank (= **EDAS**/Electronic Dictionary Administration System) wurde intern entwickelt.
- Seit zwei Jahren ist mit **Vernetziko** ein Vernetzungstool hinzugekommen.
- *elexiko* besitzt kein lexikografisches Redaktionssystem im eigentlichen Sinn

Phase der Datenverarbeitung

- Erstellung einer **Stichwortkandidatenliste** aus dem Wörterbuchkorpus und hieraus Erstellung der *elexiko*-Stichwortliste.
- Ermittlung von **Frequenzangaben** und Festlegung von Frequenzschichten, die der Definition von Bearbeitungsteilwortschätze dienen.
- Spezifikation der **Datenbankstruktur** auf Basis der DTD-Vorgaben.
- usw.

Phase der Datenanalyse

- Einsatz **automatischer Verfahren** der Datengewinnung
z. B. für Angaben zur Silbentrennung
- Einsatz automatischer Verfahren im Rahmen der
redaktionellen Arbeit, z. B. Ermittlung des Kookkur-
renzprofils eines Stichwortes
- **Verfassen der Wortartikel**, dabei auch Anlage und
Pflege von Vernetzungen sowie Integration multi-
medialer Elemente
- usw.

Phase der Vorbereitung für Online-Release

- inhaltliches und formales **Korrekturlesen**
- **Testen** der Online-Präsentation und der Recherche-möglichkeiten
- Aufbau der **Benutzungshinweise** und des Glossars
- usw.

Einige Besonderheiten im *elexiko*-Prozess

- Die Arbeit am Projekt wurde **ohne ausreichende Kenntnis** lexikografischer Prozesse und insbesondere computerlexikografischer Prozesse begonnen.
- Wichtige **Schritte** wurden so **versäumt**, z. B. Erarbeitung eines Organisationsplans (Workflow, Zeitplan, Personal) oder gründliche Marktanalyse damals schon bestehender lexikografischer Redaktionssysteme.

Einige Besonderheiten im *elexiko*-Prozess

- Der Prozess für *elexiko* musste sich in die **Gegebenheiten am IDS** einfügen, was z.B. Auswirkungen auf die Inhalte (z. B. grammatische Angaben konform mit der IDS-Grammatik), aber auch auf die Technik hatte (z. B. Einsatz von ORACLE).
- Generell gilt, dass die eben beschriebenen **Phasen** nicht hintereinander, sondern häufig **parallel** abgelaufen sind bzw. ablaufen. Der **Prozess** ist weniger linear, als in gewisser Weise **zirkulär**.

Einige Besonderheiten im *elexiko*-Prozess

- Beispielsweise ist ein Bearbeitungsteilwortschatz (Wortbildungsmittel) noch in der Planungsphase und ein anderer ist in der Phase der Datenanalyse („Lexikon zum öffentlichen Sprachgebrauch“).
- Selbst ein und dasselbe **Stichwort** kann sich **in verschiedenen Phasen** des Prozesses befinden: es kann z. B. mit automatischen Angaben schon online sein, für die nachgelagerte redaktionelle Bearbeitung setzt aber erst die Phase der Datenanalyse ein.

Vielen Dank für die Aufmerksamkeit!