

Projekt Altägyptische Wörterbücher im Verbund: Digital unterstützte Analyse der Entwicklung ägyptischer Wörterbücher

In der Frühzeit der wissenschaftlichen Untersuchung des Ägyptischen im 19. und frühen 20. Jahrhundert ist eine Vielzahl von Wörterbüchern, Wortlisten und Glossaren entstanden. Da die Ägyptische Sprache erst allmählich entschlüsselt und verstanden wurde, entwickelte sich auch die Sicht auf die Erfassung der ägyptischen Wörter weiter. Alle aktuell bekannten Wörter des Ägyptischen sind im Thesaurus Linguae Aegyptiae (<http://aaew.bbaw.de/tla>, auch Berliner Wortliste, kurz BWL) erfasst. Um nachzuvollziehen, wie sich das Verständnis des Ägyptischen über die Zeit verändert hat, schafft das vorliegende Projekt eine Infrastruktur dafür, das Vorkommen von Wörtern in ägyptischen Wörterbüchern und anderen lexikographisch relevanten Publikationen mit den Einträgen der BWL zu verknüpfen. Mit dem entwickelten Werkzeug sind automatisierte Auswertungen der Entwicklung des Verständnisses der ägyptischen Lexik – und damit der Lexikographie – möglich, die ohne IT-Unterstützung nicht denkbar wären.

Das Projekt „Altägyptische Wörterbücher im Verbund“ orientiert sich an Wörterbuchportalen wie "Wörterbuchnetz" (<http://woerterbuchnetz.de/>), "OWID" (<http://www.owid.de/>), "Etymologiebank" (<http://www.etymologiebank.nl>) und Infolux (<http://infolux.uni.lu/worterbucher/>). Spezifisch für das Ägyptische ist, dass die meisten Wörterbücher handgeschrieben sind und neben deutschen, französischen, englischen oder italienischen Textteilen auch Hieroglyphen und Partien in Demotisch, Koptisch, Griechisch, Lateinisch, Hebräisch und Arabisch enthalten (können), ganz abgesehen von Transliterationen mit Zeichen, die nicht alle im Unicode-Format existieren. Unter diesen Umständen ist eine Texterfassung per OCR und eine anschließende Auszeichnung in TEI/XML nicht möglich, so dass Verknüpfungen von Bilddaten mit einer Lemmaliste vorgenommen und Metadaten an diese Verknüpfungen angehängt werden müssen.

Die unterschiedlich strukturierten (teils nach hieroglyphischer Orthographie, teils nach Transliteration, teils nach konventionalisierter hieroglyphischer Zeichenliste) und in der Methodik der Worttransliteration individuell verfahrenen Wörterbücher machen es unmöglich und unnötig, für jedes Wörterbuch eine eigene Lemma-Liste zu erstellen und diese mit einer Hyper-Lemma-Liste zu verknüpfen. Stattdessen wird die BWL als standardisierte Lemma-Liste eingeschaltet, was zudem eine zukünftige Verknüpfung mit der ägyptischen Textdatenbank ermöglicht. So werden die alten Wörterbücher langfristig gesehen mit einem System, das ägyptische Volltexte erfasst, verschränkt.

Zur Erfassung und Aufbereitung der Informationen über die Entwicklung ägyptischer Wörterbücher wurde ein webbasiertes Werkzeug erstellt. Dieses bietet eine für mehrere Forscher mit differenzierenden Fragestellungen gleichzeitig zugängliche Oberfläche, um die Verknüpfung von Wörterbuch- und anderen Publikationseinträgen mit Wörtern der BWL zu ermöglichen. Die BWL wurde in einem Vorverarbeitungsschritt extrahiert, so dass die Wörter mit Metadaten als Graphiken vorlagen. Die einzelnen Seiten der oft handgeschriebenen Publikation lagen ebenfalls als Graphiken vor. Metadaten zu den Graphiken und zur Verknüpfung wurden in einer MySQL-Datenbank gespeichert, auf die anschließend mit Hibernate zugegriffen wurde. Hierbei wurde bspw. die Möglichkeit geschaffen, zu jeder einzelnen Verknüpfung zu annotieren, wie zutreffend die damalige Transliteration und Übersetzung aus heutiger Sicht waren.

Um die Verknüpfung herzustellen, musste die Möglichkeit geschaffen werden, in einer Eingabeoberfläche zu erfassen, welcher Bereich einer Seite mit welchem Eintrag der BWL zusammenhängt. JSF und PrimeFaces enthalten bereits verschiedene Funktionalitäten, um diese Verknüpfung herzustellen, bspw. Datentabellen mit Suchfunktion für die Wörter der BWL und ein Verfahren, um Bildbereiche in den Publikationsseiten zu extrahieren. Aus diesem Grund wurden diese zur Umsetzung der Weboberfläche eingesetzt und so ein Werkzeug geschaffen, mit dem sich Wörter der BWL mit Publikationsseitenbereichen verknüpfen lassen.

Weiterhin wurde eine Möglichkeit geschaffen, die Publikationen nach festgelegten Kriterien zu analysieren. Hierbei wurden Anfrageoptionen nach Metadaten der einzelnen Verknüpfungen und statistischen Daten der Metadaten mit MySQL implementiert. Diese lassen sich bei Bedarf problemlos erweitern. Auf diese Weise lassen sich die großen Datenmengen der Wörterbücher schnell und effizient auswerten.

Mit dem beschriebenen Werkzeug wurden seitdem verschiedene ägyptische Wörterbücher des 19. Jahrhunderts erfasst und analysiert. Hierbei wurden in insgesamt zehn Publikationen 30371 Verknüpfungen erstellt. Die sukzessive erweiterbare Datenbank und die installierten Funktionen werden den Nutzer in die Lage versetzen, die einzelnen Etappen der Erforschung der ägyptischen Lexik und die jeweils zu Grunde gelegte Konzeption und Methodik besser nachvollziehen zu können. Im idealen Falle soll es möglich sein, die Forschungsgeschichte eines Lexems von den Anfängen der Ägyptologie bis zum Erscheinen des Wörterbuchs der Ägyptischen Sprache (1926-31), das die Basis der BWL darstellt, und darüber hinaus skizzieren zu können. Je nach Suchkriterium können dabei exemplarische bzw. statistische Daten ausgewertet werden, wie z.B. Qualität und Quantität der verwendeten Primär- und Sekundärquellen, Häufigkeit von lexikographischen Fehlern und Missdeutungen, die entweder mangels besserer Kenntnis und auch gelegentlich aufgrund fehlender Sorgfalt vorgekommen sind.

Insgesamt ermöglicht die so geschaffene IT-Infrastruktur eine Analyse der frühen ägyptischen Lexikographie, die sonst nicht bzw. nur durch sehr zeitaufwändige Recherchen möglich wäre. Die Präsentation wird sowohl die zugrundeliegenden Fragestellungen der Ägyptologie und bereits gefundene Lösungen als auch die technische Infrastruktur, die zur Unterstützung der Beantwortung der Fragestellungen geschaffen wurde, darstellen.