

URN: urn:nbn:de:kobv:b4-opus-24368

INGELORE HAFEMANN & PETER DILS,
Der Thesaurus Linguae Aegyptiae – Konzepte und Perspektiven,

in:

Perspektiven einer corpusbasierten historischen Linguistik und Philologie. Internationale Tagung des Akademienvorhabens „Altägyptisches Wörterbuch“ an der Berlin-Brandenburgischen Akademie der Wissenschaften, 12. – 13. Dezember 2011, herausgegeben von Ingelore Hafemann, Berlin 2013, S. 127-143.

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Thesaurus Linguae Aegyptiae 4

Perspektiven einer corpusbasierten historischen Linguistik und
Philologie. Internationale Tagung des Akademienvorhabens
„Altägyptisches Wörterbuch“ an der Berlin-Brandenburgischen
Akademie der Wissenschaften, 12. – 13. Dezember 2011

herausgegeben von Ingelore Hafemann

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Thesaurus Linguae Aegyptiae

4

BERLIN 2013

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Perspektiven einer corpusbasierten historischen Linguistik
und Philologie

Internationale Tagung des Akademienvorhabens „Altägyptisches
Wörterbuch“ an der Berlin-Brandenburgischen Akademie der
Wissenschaften, 12. – 13. Dezember 2011

herausgegeben von Ingelore Hafemann

BERLIN

2013

Dieser Band wurde durch die gemeinsame Wissenschaftskonferenz im Akademienprogramm mit Mitteln des Bundes (Bundesministerium für Bildung und Forschung) und des Landes Berlin (Senatsverwaltung für Wirtschaft, Technologie und Forschung) gefördert

Die Publikation unterliegt folgender Creative-Commons-Lizenz:
„Namensnennung – Keine kommerzielle Nutzung – Weitergabe unter
gleichen Bedingungen 3.0 Deutschland“

<http://creativecommons.org/licenses/by-nc-sa/3.0/de/>



URN: urn:nbn:de:kobv:b4-opus-24310

INHALTSVERZEICHNIS

| | |
|--|-----|
| VORWORT | 7 |
| GREGORY CRANE & ALISON BABEU Global Editions and the Dialogue among Civilizations | 11 |
| HISTORISCHE CORPUS-PROJEKTE – SYNCHRON UND DIACHRON | |
| STÉPHANE POLIS & JEAN WINAND The Ramses project. Methodology and practices in the annotation of Late Egyptian Texts | 81 |
| SERGE ROSMORDUC The Ramses project in perspective. Managing evolving linguistic data | 109 |
| DIETER KURTH Das Edfu-Projekt. Ziel, Methode und Verarbeitung der lexikographischen Ergebnisse | 121 |
| INGELORE HAFEMANN & PETER DILS Der Thesaurus Linguae Aegyptiae – Konzepte und Perspektiven | 127 |
| GÜNTER VITTMANN Zur Arbeit an der Demotischen Textdatenbank: Textauswahl | 145 |
| GERNOT WILHELM Das Hethitologie Portal Mainz | 155 |
| JOST GIPPERT The TITUS Project. 25 years of corpus building in ancient languages | 169 |
| KURT GÄRTNER & RALF PLATE Die Doppelfunktion des digitalen Textarchivs als Wörterbuchbasis und als Komponente der Online-Publikation. Am Beispiel des Mittelhochdeutschen Wörterbuchs | 193 |
| HANS-CHRISTIAN SCHMITZ, BERNHARD SCHRÖDER & KLAUS-PETER WEGERA Das Bonner Frühneuhochdeutsch-Korpus und das Referenzkorpus ,Frühneuhochdeutsch‘ | 205 |

| | |
|--|-----|
| ALEXANDER GEYKEN Wege zu einem historischen Referenzkorpus des Deutschen: das Projekt Deutsches Textarchiv | 221 |
| BRYAN JURISH Canonicalizing the Deutsches Textarchiv | 235 |
| WORTGESCHICHTE - TEXTGESCHICHTE - SPRACHGESCHICHTE: TRADITION UND INNOVATION BEI DER TEXTPRODUKTION | |
| FRANK FEDER & SIMON D. SCHWEITZER Auf dem Weg zu einem integrierten Lexikon des Ägyptisch- Koptischen | 245 |
| FRIEDHELM HOFFMANN Die Demotische Wortliste – virtuell erweitert | 263 |
| GÜNTER VITTMANN Kursivhieratische Texte aus sprachlicher und onomastischer Sicht | 269 |
| MATHEW ALMOND, JOOST HAGEN, KATRIN JOHN, TONIO SEBASTIAN RICHTER & VINCENT WALTER Kontaktinduzierter Sprachwandel des Ägyptisch-Koptischen: Lehnwort-Lexikographie im Projekt Database and Dictionary of Greek Loanwords in Coptic (DDGLC) | 283 |
| THOMAS GLONING Historischer Wortgebrauch und Themengeschichte. Grundfragen, Corpora, Dokumentationsformen | 317 |
| LOUISE GESTERMANN Die altägyptischen Sargtexte in diachroner Überlieferung | 371 |
| THOMAS STÄDTLER Überlegungen zu Textsorte und Diskurstradition bei der Beschreibung von Textcorpora und ihr Bezug zur lexikographischen Forschung | 385 |

VORWORT

Die internationale Tagung „Perspektiven einer corpusbasierten historischen Linguistik und Philologie“ vom 12. – 13. Dezember 2011 am Akademienvorhaben „Altägyptisches Wörterbuch“ der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) war dem Thema des Aufbaus und der Nutzungsperspektiven elektronischer Textcorpora und Wörterbücher in den historischen Sprachen gewidmet. Die Teilnehmer, Vertreter der Ägyptologie, der Hethitologie, Indogermanistik sowie Referenten aus der historischen Lexikographie des Mittel- und Frühneuhochdeutschen und des Altfranzösischen diskutierten vor allem über die Veränderungen, die mit dem Einsatz elektronischer Erfassungs- und Verarbeitungsprozeduren einhergehen. Vertreter der Computerlinguistik vom „Zentrum Sprache“ der BBAW wurden in die Diskussionen einbezogen. Dort beschäftigt man sich seit Jahren mit dem Aufbau großer elektronischer Textcorpora (DWDS), darunter auch solcher, die historische Texte (DTA) für die elektronische Nutzung ermöglichen.

Die größte Herausforderung dieser neuen elektronischen Corpora und Wörterbücher ist es, sowohl den Methoden und damit den wissenschaftlichen Ansprüchen der traditionellen Philologie und Lexikographie unbedingt verpflichtet zu bleiben als auch neue Gebiete wie die Corpus- und Computerlinguistik für die historischen Sprachen zu öffnen. Die Teilnehmer haben gemeinsam und disziplinenübergreifend die Möglichkeiten und Grenzen der Datenerfassung, ihrer Präsentation und den Nutzen neuer Auswertungsprozeduren diskutiert.

Unter dem ersten Thema „Historische Corpusprojekte – synchron und diachron“ wurden elektronische Corpora vorgestellt und ein intensiver Austausch darüber geführt, welche Datenstrukturen die linguistischen Inhalte in adäquater Weise abbilden. Wichtig war die Frage, auf welche Resonanz diese elektronischen Corpora bei den Nutzern gestoßen sind und welche Erwartungen und Anforderungen aus den verschiedenen Fachdisziplinen an die Projekte herangetragen werden. Der Austausch über Nutzungsperspektiven elektronischer Corpora schloss auch die Diskussion über die Erarbeitung projektübergreifend einsetzbarer Standards der Codierung und Strukturierung historischer Textdaten mit ein. Hinsichtlich einer mittel- und langfristigen Nutzbarkeit sowie einer langfristigen Datensicherheit stehen solche Fragen zunehmend im Focus und einige aktuelle Initiativen dazu wurden vorgestellt. Spezielle technische Aspekte

elektronischer Datenerfassung und automatischer Analyse- und Speicherungsverfahren elektronischer Textdaten konnten am letzten Tag als ein Themenschwerpunkt mit den Programmierern diskutiert werden.

Ein zweiter Schwerpunkt waren konkrete Fragstellungen aus der historischen Lexikographie und diachronen Textanalyse. Für das Ägyptische ist der diachrone Ansatz auf Grund der über vier-tausendjährigen Textüberlieferung von großer Relevanz. Themen wie historischer und/oder textgattungsspezifischer Wortgebrauch, die Erarbeitung diachroner Wortlisten und Aspekte des kontaktindizierten Sprachwandels konnten disziplinübergreifend zwischen den Ägyptologen und den Kollegen der historischen Lexikographie des Mittel- und Frühneuhochdeutschen und des Altfranzösischen behandelt werden.

Mit dem Abendreferenten Gregory Crane, dem Begründer der „Perseus Digital Library“, wurde ein breites Publikum angesprochen. In seinem Vortrag hat er noch einmal die hohe Relevanz und die neuen Möglichkeiten der Einbeziehung zahlreicher Wissenschaftler und einer interessierten Öffentlichkeit in die Projektarbeit demonstriert, die das Internet auf völlig neue Weise eröffnet hat. Die Herausgeberin ist sehr froh, seinen programmatischen Beitrag zu diesem Thema, dessen schriftliche Form er gemeinsam mit Alison Babeu erarbeitet hat, ebenfalls in diesem Band präsentieren zu können.

Wir danken der Berlin-Brandenburgischen Akademie der Wissenschaften für die umfassende Unterstützung unserer Projektarbeit und ganz speziell der Vorbereitung dieser Konferenz sowie der Möglichkeit, die Akten auf dem E-Doc-Server der Akademie veröffentlichen zu können.

Der Hermann und Elise geborene Heckmann Wentzel-Stiftung sei hiermit ausdrücklich für die unbürokratische und großzügige finanzielle Unterstützung dieser erfolgreichen Tagung gedankt.

Das Akademienvorhaben „Altägyptisches Wörterbuch“ konnte sich als aktives Mitglied des Weiteren auf das „Zentrum Grundlagenforschung Alte Welt“ stützen, dem alle altertumswissenschaftlichen Vorhaben der BBAW angehören. Dem Zentrum ist es zu danken, dass der Abendvortrag von Gregory Crane einem breiteren Publikum dargeboten werden konnte.

Allen Autoren dankt die Herausgeberin für ihre anregenden Diskussionen und die qualitätvollen Beiträge in diesem Band.

Auf eine Gesamtbibliographie wurde verzichtet und die Abkürzungen der in den ägyptologischen Beiträgen erwähnten Zeitschriften und Reihen folgen dem Lexikon der Ägyptologie, herausgegeben von Wolfgang Helck und Wolfhart Westendorf, Band VII: Nachträge, Korrekturen, Indices, Wiesbaden 1992, XIV-XIX.

Ganz besonders sei schließlich Frau Angela Böhme für die gewissenhafte redaktionelle Bearbeitung der Manuskripte gedankt sowie Dr. Simon Schweitzer für seine Hilfe beim Erstellen des Layouts.

Berlin, Mai 2013

Ingelore Hafemann

DER THESAURUS LINGVAE AEGYPTIAE – KONZEPTE UND PERSPEKTIVEN

INGELORE HAFEMANN & PETER DILS

Teil I: Einführung, Grundstrukturen und Nutzung (Ingelore Hafemann)

Zu den interessanten Internetangeboten auf dem Gebiet der historischen Textwissenschaft und Lexikographie gehört seit einigen Jahren zweifellos auch der *Thesaurus Linguae Aegyptiae* (kurz TLA). Er umfasst ein digitales Corpus von Abschriften ägyptischer Texte aus der Pharaonenzeit (von ca. 2400 v. Chr. bis zum 5. Jh. n. Chr.). Bei den Texten handelt es sich um Dokumente aus der gesamten Geschichte der Pharaonenherrschaft, die nicht nur in unterschiedlichen Sprachstufen des Ägyptischen verfasst, sondern auch in verschiedenen Schriften niedergeschrieben wurden – in hieroglyphischer und hieratischer sowie in demotischer Schrift. Das Corpus bietet die Texte komplett in ägyptologischer Transkription und mit Übersetzungen und Kommentaren an. Teils werden auch Fotos und Umzeichnungen der Texte gegeben. Die Volltexte stehen für lexikalische und philologische Recherchen im Internet zur Verfügung. Hauptnutzer sind dabei Ägyptologen, aber dank der Übersetzungen können auch Vertreter verschiedenster Nachbardisziplinen sowie Kulturwissenschaftler oder interessierte Laien dieses Corpus mit Gewinn einsehen. Konkret stellt dieser Thesaurus die Publikationsplattform dreier Akademienvorhaben dar.¹ Das sind einmal die beiden Vorhaben mit dem Namen *Altägyptisches Wörterbuch*, eines an der Berlin-Brandenburgischen Akademie der Wissenschaften und eines an der Sächsischen Akademie der Wissenschaften zu Leipzig. Das dritte Partnerprojekt ist die *Demotische Textdatenbank* an der Akademie der Wissenschaften und der Literatur, Mainz. Der Zusammenarbeit dieser drei Akademienvorhaben haben sich weitere Projektgruppen angeschlossen, wie das Projekt *Leuven online index of Ptolemaic and Roman Hieroglyphic Texts* der Katholieke Universiteit Leuven und das *Bonner Totenbuchprojekt*, das an der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste verankert ist, sowie das abgeschlossene Corpusprojekt *Digital Heka* der Universität Leipzig, das sich der Erfassung magischer Sprüche und Zaubertexte widmete.

¹ URL: <http://aew.bbaw.de/tla/>.

Konzepte

Der *Thesaurus Linguae Aegyptiae* ist eine lexikalische Datenbank ägyptischer Texte. Durch die Verknüpfung einer Textdatenbank mit einem lexikalischen Wort-Thesaurus innerhalb eines modernen Navigationsprogramms sind vielfältige Abfragemöglichkeiten geschaffen worden, die weit über gedruckte Formate von Texteditionen und Wörterbüchern hinausgehen, diese aber nicht ersetzen sollen und können. Der TLA ist ein völlig neuartiges Instrument der Recherche und Forschung, das im Folgenden beschrieben wird.

Das Konzept eines solchen digitalen Textcorpus muss im Rahmen dieses Tagungsbandes nicht erläutert werden. Inzwischen ist der TLA auch in dem kleinen Fach der Ägyptologie als ein stabiles Forschungswerkzeug etabliert. Allerdings ist das Gesamtkonzept eines digitalen Corpus mit allen seinen Nutzungsstrategien bisher nur wenigen Ägyptologen bekannt, da die Arbeit mit digitalen Corpora in der Ägyptologie wie in den meisten historischen Sprachwissenschaften noch Neuland ist.

Die Arbeit am elektronischen Corpus – die 1992 gestartet wurde – ist von Beginn an einem Grundprinzip verpflichtet gewesen: dem der corpusbasierten Lexikographie. Im Fokus stand demnach zwar ein Text-Thesaurus – das Corpus – aber in erster Linie ist das Neuprojekt als ein lexikographisches Unternehmen anzusehen. Dieses folgte dem methodischen Grundprinzip des alten Vorgängerprojektes an der damals Berliner Akademie, das im Jahre 1897 von Adolf Erman als das internationale Großprojekt „Wörterbuch der ägyptischen Sprache“ an der Preußischen Akademie initiiert wurde. In diesem Projekt verfolgte Erman konsequent den Ansatz der Darstellung des Wortschatzes auf der Grundlage der kompletten Quelltexte. Sein großes Wörterbuch² ist bis heute das Standardwerk zur ägyptischen Lexikographie. Das eigens dafür angelegte Archiv mit Textabschriften, auf über 1,5 Millionen lexikalisch-lexikographisch sortierten Zetteln niedergelegt – das sogenannte „Zettelarchiv des ägyptischen Wörterbuchs“ – ist neben dem Wörterbuch ebenfalls ein Forschungswerkzeug ersten Ranges für Ägyptologen geblieben.

Eben diesem Prinzip der textbasierten Arbeitsweise folgte auch das 1992 an der Berlin-Brandenburgischen Akademie der Wissenschaften neu gegründete Projekt „Altägyptisches Wörterbuch“. Ein elektronisches Textcorpus mit einer integrierten elektronischen Wort-

² ERMAN, A. & H. GRAPOW (Hrsg.), *Wörterbuch der aegyptischen Sprache*, 7 Bde., Leipzig / Berlin 1926-1963, Belegstellen, 5 Bde., Leipzig / Berlin 1935-1953.

liste soll ein völlig neuartiges Forschungswerkzeug zur Verfügung stellen, das nun Wörterbuch und Textarchiv miteinander verbindet und dem Nutzer komplexe Recherchen ermöglicht. Erstmals stehen ihm in diesem System nun alle ausgewerteten Quellentexte zur Verfügung.

Strukturen

Die erwähnte elektronische Wortliste (Lemmaliste) ist das lexikalische Rückgrat der Struktur. Diese Lemmaliste hat zweierlei Funktionen: zum einen ist sie ein Nachschlagewerk zum Wortschatz des Ägyptischen für jeden Nutzer, zum anderen dient sie intern als Lemmatisierungswerkzeug innerhalb des Programms der Texterfassung. Sie wurde mit dem bis heute bekannten Wortschatz des Ägyptischen entsprechend dem aktuellen Forschungsstand angefüllt.³ Hierzu wurden lexikographische Standardwerke ausgewertet sowie mit neuen Wörtern aus den eingegebenen Einzeltexten vervollständigt.

Die lexikographische Beschreibung der einzelnen Lemmata in der Liste ist relativ flach und setzt sich aus den Angaben zur Transkription, hieroglyphischen Schreibweise, Bedeutungsangabe in Form von Übersetzungen (Deutsch und Englisch), der Wortkategorie und den bibliographischen Referenzen einiger lexikographischer Standardwerke zusammen (vgl. Abb. 1). Den Wörtern des Spezialwortschatzes sind Sachgruppenbezeichnungen zugewiesen worden, wie Toponym, Königs- und Göttername, Personennamen, Titel und Epitheton sowie Eigenname von Sachen und Institutionen, die in der Rubrik Wortkategorie stehen. Es handelt sich bei diesen Wörtern immer um Substantive und insbesondere oft um Komposita.

| Angaben zum Lemma  rmi (Lemma-Nummer 94180) | |
|--|---|
| Übersetzung | weinen; beweinen (to weep; to bewweep) |
| Kurzreferenz | Wb 2, 416-417.10; FCD 149 |
| Wortkategorie | Vb., 3ae inf. |

Abb. 1: Detailansicht zum Lemma *rmi* im TLA

³ Ende 2012 zählte die ägyptische Lemmaliste 31.400 Lemmata, inklusive der 3.000 Personennamen, und die demotische Lemmaliste 11.400 Lemmata, einschließlich der 1.800 Personennamen.

Die Angaben zum Gebrauch der Wörter und zu ihren semantischen Lesarten sowie zur phraseologischen und idiomatischen Verwendung, die sonst in Wörterbüchern verzeichnet werden, sind nur teilweise als Einträge in die Lemmaliste aufgenommen. Der vielfältige Wortgebrauch aber kann aus dem elektronischen Corpus selbst entnommen werden, denn erst im Sprachgebrauch in den Texten entwickeln sich die lexikalischen und grammatischen Potenzen der Wörter, zeigt sich ihre Polysemie und werden Metonyme sichtbar oder auch die dialektale Färbung von Wortbedeutungen und sogar Ideosynkrasien einzelner Autoren.

Daher lag und liegt die tatsächliche Hauptarbeit des Projektes in der Schaffung eines Corpus der Quellentexte. Die Texteingabe erfolgt ausschließlich manuell. Unmittelbar bei der Eingabe der einzelnen Textwörter wird in einem eigenen Arbeitsschritt Wort für Wort mit der Lemmaliste verlinkt – also lemmatisiert. Diese Lemmazuweisung erfolgt halbautomatisch und muss jeweils manuell bestätigt werden. Damit werden gleichzeitig die lexikographischen Beschreibungslabel des Lemmas mit jedem Textwort verknüpft, wie z.B. die Wortkategorie. Danach werden grammatisch-morphologische Annotationen, wie Flexionsform, für die aktuellen Wortformen jedes Textwortes vorgenommen⁴ und es wird eine gebundene Satzübersetzung angefertigt.

Den Texten selbst werden auch Metadaten zugewiesen wie Datierung, Textsorte, Herkunft, Textträger oder Schriftform. Diese Metadaten werden als Filter für Abfragen dienen und dann können bei den Wortrecherchen nur bestimmte Texte einer bestimmten Zeit oder Herkunft einbezogen werden. Hier wird jeder Nutzer frei sein und kann seine Fragen nach eigenen Bedürfnissen formulieren.

Was wir gewinnen, ist ein vollständig corpusbasiertes Wörterverzeichnis – oder aber ein lexikonbasiertes Textcorpus, je nachdem, aus welcher Perspektive man es betrachtet. Im Laufe der Arbeit am Corpus ergab sich tatsächlich auch ein Perspektivwechsel vom Lexikon zum Corpus, und die wissenschaftliche Dimension eines elektronischen Corpus als solchem wurde immer deutlicher. Damit gerieten die Methoden der Corpuslinguistik in den Blick, einer neuen

⁴ Es sind nicht alle Texte grammatisch annotiert worden; das gesamte Corpus der Demotischen Textdatenbank ist nicht grammatisch annotiert. Die grammatische Annotation ist bis Ende 2012 noch nicht im TLA abrufbar.

Disziplin, die mit dem Aufbau elektronischer Corpora neue Wege in der Linguistik aufgetan hat.⁵

Leistungen und Nutzungsstrategien des TLA

Ursprünglich stand also die Idee einer reinen Belegstellenabfrage für die Wörter des ägyptischen Lexikons im Mittelpunkt. Da das gesamte Textcorpus Wort für Wort lexikalisch verknüpft ist, werden den Lemmata der elektronischen Wortliste sukzessive mit der laufenden Texterfassung immer weitere Belege zugewiesen. Mit dem Corpus wächst die Belegmenge je Lemma.

So wurde das Konzept eines virtuellen Wörterbuchs entwickelt. Die Philosophie dahinter ist die folgende: Der durch den Lexikographen erarbeitete traditionelle Wörterbucheintrag, der zu den einzelnen Wörtern bestimmte Angaben abschließend in einem Buch oder auch einem digitalen Dokument verzeichnet, wird hier durch ein Werkzeug ergänzt, das dem Forscher erlaubt, die gesamte Materialbasis zur Aufklärung seiner spezifischen Wortrecherche zu nutzen. Bekanntlich ist jeder Eintrag in gedruckten Wörterbüchern eine Notlösung. Es lassen sich meist viele weitere Wortbedeutungen oder Lesarten eines Lemmas bzw. semantische Eigenschaften durch kontextuelle Zusammenhänge im tatsächlichen Sprachgebrauch finden. Die Fragen des Forschers lassen sich mit einem lexikonbasierten Corpus frei und immer wieder neu formulieren. Die angebotenen Funktionen des TLA sind vornehmlich auf ägyptologische Bedürfnisse abgestimmt. Es können Formenbildung von Verben und Substantiven sowie Vorkommen bestimmter Phonemfolgen ermittelt werden. Neben den wortspezifischen Abfragen können auch Recherchen zu phraseologischen und idiomatischen Wendungen oder zum Wortgebrauch im Gesamtkorpus oder in bestimmten Textgruppen angestellt werden. Durch den Zugang zum Gesamtkorpus lassen sich allgemeine Fragen der Textlinguistik, die sich mit satzübergreifenden Strukturen beschäftigen, erstmals am ägyptischen Textmaterial durchführen. Dazu gehören Fragen nach der Lexemvariation eines Textes oder seiner Kohärenz. Auch thematische und funktionale Aspekte der Textstruktur sind auf neuer Datengrundlage untersuchbar. Einige Analysen, die aus dem Bereich der Corpuslinguistik bekannt sind, wie Wortstatistiken, Verteilung von Worthäufigkeiten,

⁵ LEMNITZER, L. & H. ZINSMEISTER, *Korpuslinguistik*, 2. Aufl., Tübingen 2012; LÜDELING, A. & M. KYTÖ, *Corpus linguistics: An International Handbook*, Berlin / New York, Vol. 1 2008, Vol. 2 2009.

Schlüsselwortanalysen und Kollokationsanalysen, sind bereits im TLA lauffähig implementiert und können nach Eingabe der gewünschten Parameter per Mausklick gestartet werden (s. detaillierter im Teil II unten). Mit Hilfetexten in Deutsch und Englisch und einem Handbuch (als PDF) soll allen Nutzern der multifunktionale Zugang zum TLA erleichtert werden.

Seit 2009 kann der Nutzer auch nach hieroglyphischen Schreibungen innerhalb der Lemmata der Wortliste recherchieren, wobei jeweils eine Standardschreibung je Lemma geboten wird.

Lemmata anhand hieroglyphischer Schreibungen suchen

Bevor Sie diese Funktion nutzen, lesen Sie unbedingt die [Informationen zum Datenbestand](#) und die [Regeln zur Codierung der Suchanfrage](#)

Zeichen Codes

suchen nach

[Zeichenliste](#)

Abb. 2a: Suche nach hieroglyphischen Schreibungen im TLA

Gerätschaften aus Haus und Grab

| | | | | | | | | | |
|---|---|---|---|---|---|--|--|---|---|
|  Q1 |  Q2 |  Q3 |  Q4 |  Q5 |  Q6 |  Q7 |  Q7A |  Q11 |  Q12 |
|  Q13 |  Q14 |  Q16 |  Q24 |  Q29B |  Q36A |  Q37 |  Q45 |  Q100 |  Q101 |
|  Q102 |  Q103 |  Q104 |  Q105 |  Q106 |  Q107 | | | | |

[Zeichenliste](#)

Abb. 2b: Ausschnitt aus der Zeichenliste

Thesaurus Linguae Aegyptiae

Lemmata - Ergebnis der Suche nach hieroglyphischen Schreibungen
Suche wurde ausgeführt anhand von hieroglyphischen Graphemen

173 passende Schreibungen gefunden
[Anfang] [zurück] [weiter] [Ende]

| Schreibung | Lemma | nähere Angaben zur Schreibung | bibliographische Referenz und Urheber |
|--|---|--|--|
|  | .As.t "Isis" (Wb 1, 20; 4, 8.11-13; LGG I, 61 ff.) (Lemma-Nummer 271) | (Ansetzungsform) Standardschreibung | s. Literaturangaben zum Lemma (AAeW / BBAW) |
|  | .As.t "Aset, Isis" (RPN I 3.18) (Lemma-Nummer 400357) | (Ansetzungsform) Standardschreibung | s. Literaturangaben zum Lemma (AAeW / BBAW) |
|  | .As.tj (?) "PNj?" (RPN I 4.12) (Lemma-Nummer 709479) | (Ansetzungsform) Standardschreibung | s. Literaturangaben zum Lemma (AAeW / BBAW) |
|  | .As.tj "PNj?" (RPN I 295.8) (Lemma-Nummer 706340) | (Ansetzungsform) Standardschreibung | s. Literaturangaben zum Lemma (AAeW / BBAW) |
|  | .As.tj-r-s(θ) "PNj?" (RPN I 3.19) (Lemma-Nummer 706473) | (Ansetzungsform) Standardschreibung | s. Literaturangaben zum Lemma (AAeW / BBAW) |
|  | .As.tj-b-m-nr.t-p.t "PNj?" (RPN I 298.21) (Lemma-Nummer 706429) | (Ansetzungsform) Standardschreibung | s. Literaturangaben zum Lemma (AAeW / BBAW) |
|  | .As.tj-r-s "PNj?" (RPN I 3.20) (Lemma-Nummer 706475) | (Ansetzungsform) Standardschreibung | s. Literaturangaben zum Lemma (AAeW / BBAW) |

Abb. 2c: Ergebnisliste der Suche nach der Hieroglyphe Q1

Seit 2011 kann man hieroglyphische Schreibungen auch für einige Textwörter des Corpus einsehen, vorerst probeweise in einem ausgewählten Teilcorpus. Dieser Ansatz wird ausgebaut und eröffnet neue Perspektiven für die Erforschung von Verteilungen und Frequenzen hieroglyphischer Zeichen in den Wortschreibungen, die dann nach linguistischen, soziolinguistischen, chronologischen oder regionalen Parametern angeordnet und untersucht werden können.

Teil II: Funktionalitäten und Anwendungen des TLA (Peter Dils)

Im Bereich der Funktionalität bietet der TLA zwei Einstiegsmöglichkeiten in die digitalen Daten. Die erste Möglichkeit läuft über die einzelnen Lemmata der hieroglyphisch-hieratischen oder alternativ die der demotischen Lemmaliste. Die zweite startet über die Texte des Textcorpus.

Die erste Einstiegsmöglichkeit über die Lemmaliste dient vorrangig ausgebildeten Ägyptologen, wobei die Belegstellenabfrage im Vordergrund steht. Die zweite Möglichkeit des Einstiegs in das Text-

corpus oder auch nur in Teilcorpora wird gern von Studierenden genutzt – besonders für das Corpus der literarischen Texte, die häufig Gegenstand von Seminaren zur Textlektüre sind. Dabei werden die Transkriptionen, Übersetzungen und Kommentierungen als Vorbereitungsgrundlage für den Unterricht konsultiert.

Angaben zum Lemma
wnm (Lemma-Nummer 46710)

| | |
|---------------|--------------------|
| Übersetzung | essen (to eat) |
| Kurzreferenz | Wb 1, 320.1-321.12 |
| Wortkategorie | Vb., 3rad. |

für dieses Wort gibt es in der Textdatenbank 462 Belegstelle(n)
zur Anzeige der Belegstellen wählen Sie in den Feldern des Formulars Ihre Ausgabeoptionen und klicken Sie auf die Schaltfläche "Start", um den Suchvorgang zu beginnen.

| | |
|------------------------------|---|
| Belege für | <input type="text" value="nur dieses Lemma"/> |
| Ausgabeformat | <input type="text" value="satzweise"/> |
| Übersetzung der Sätze | <input type="text" value="anzeigen"/> |

Werte zurücksetzen
Start

[Suchen nach Kombinationen dieses Wortes mit einem anderen Wort](#)

[Kollokationsanalyse für dieses Wort anfordern](#)

[Analyse der lexikalischen Gravitation für dieses Wort anfordern](#)

für dieses Wort gibt es im Digitalisierten Zettelarchiv 900 Bilder

Abb. 3a: Detailansicht zum Lemma wnm im TLA

pMillingen, Die Lehre des Amenemhet [nähere Angaben zum Text] [der Text in der Objekthierarchie]

(Anzeige steht am Textanfang)

[1, 1] § (H1a) (Rubrum: HA,t-a-m sbA,yt.) VP jri.t.n Hm n.(j) nsw-bj,tj (Kartusche|sHtp-jb-ra|Kartusche) VP § (H1b) zA-ra (Kartusche|jmn-m-HA,t|Kartusche) VP mAa-xw VP § (H1c) Dd=f m [1, 2] wpi.t mAa,t VP n zA =f nb-r-Dr VP

§ (H1d) Dd=f xai m nTr VP

sDm n Dd.tj =j [n] =k VP nswy =k tA VP HqA,y =k [1, 3] jdb.(Pt.) VP § (H1e) jri =k HA, w Hr nfr VP

Anfang der (wörtl.: aus/mit der) Lehre, die verfaßt hat die Majestät des Königs von Ober- und Unterägypten (Kartusche|Sehetep-ib-Re|Kartusche), der Sohn des Re (Kartusche|Amenemhat|Kartusche), der Gerechtfertigte, der (wörtl.: indem er) in einem Offenlegen der Wahrheit (d.h. in einer Vision?) zu seinem Sohn, dem Allherm, spricht. ^(C)

Erschienen als Gott, sagt er. ^(C)

Höre auf das, was ich dir sagen werde, (damit) du das Land regieren wirst, (damit) du die Ufergebiete beherrschen wirst, (indem/und) du (dabei) einen Überschuß an Vollkommenheit erreichst. ^(C)

Abb. 3b: Beispieltext aus dem TLA

Die einzelnen transkribierten Wortformen im Text liefern den Zugang zu den passenden Lemmata in der Wortliste, über die dann alle Belegstellen dieses Wortes im Corpus aufgelistet werden, und zwar in ihrem textlichen Zusammenhang. Kontextuelle Übersetzungen pro Satz und Quellenangabe sowie chronologische und geographische Parameter liefern per Mausklick schnell einen Überblick über Wörter und ihren Gebrauch. Parallel und synchron können für jedes Lemma der Wortliste die betreffenden Zettel aus dem digitalisierten Zettelarchiv – im Beispiel 900 Zettel – von den insgesamt über 1,5 Millionen Zetteln des Wörterbucharchivs angezeigt werden, das seinerzeit – wie oben beschrieben – Adolf Erman und Hermann Grapow angelegt hatten.

Spezielle Anzeigefunktionen

Die Ergebnisse können in doppelter Hinsicht konkretisiert werden. Zum einen kann man nach dem gemeinsamen Vorkommen von zwei bestimmten Lemmata suchen, oder der Nutzer sucht nach einem Lemma und dessen Kombination mit einem Wort einer bestimmten Wortart wie zum Beispiel dem Vorkommen eines bestimmten Verbs der Bewegung mit bestimmten Präpositionen. Zum anderen kann eine Kollokationsanalyse für ein Lemma angefordert werden, so dass eine Liste von – statistisch gesehen – regelmäßigen gemeinsam auftretenden Lemmata ausgegeben wird oder auch von solchen, die unerwartet zusammen auftreten. Bei der Kollokationsanalyse wird also versucht, Nuancen des Wortgebrauchs durch Sammlung und Ordnung typischer Wortkombinationen zu erfassen, was auf der Basis von Prüfstatistiken errechnet und in Tabellenform dargestellt werden kann. Im in Abb. 4 gezeigten Beispiel wird gefragt, welche Lemmata typischerweise eine Stelle rechts vom ägyptischen Verb *wnm*: „essen“ vorkommen.⁶ Im TLA sind hierfür zwei statistische Analyseverfahren – *T-score* und *MI-score* genannt⁷ – implementiert. Nach den Regeln der ägyptischen Grammatik ist an der Stelle rechts vom Verb am ehesten entweder das Subjekt oder das Objekt zu erwarten und das wird durch die Analyse mittels *T-score* bestätigt: „ich esse, er isst, du isst, sie essen“ einerseits sowie andererseits „Brot essen, Nahrungsratio/Brot essen, Kleinvieh essen, Kot essen, ihn essen, mich essen“.

⁶ Für die Analyse sind in der Eingabemaske die Parameter für die gewünschten Wortabstände einstellbar; s. Handbuch.

⁷ *T-scores* zeigen häufige Kollokationen, *MI-scores* weisen eher seltene, aber enge Zusammenhänge terminologischen Charakters aus.

Kollokationsanalyse

node Wort: *wmm* "essen" (Wb 1, 320.1-321.12)

| | | | |
|--|---|--------------------------------|---|
| maximaler Abstand nach links | minimaler Abstand nach links | minimaler Abstand nach rechts | maximaler Abstand nach rechts |
| <input type="text" value="0"/> | <input type="text" value="0"/> | <input type="text" value="1"/> | <input type="text" value="1"/> |
| Analyse ausführen für | Kollokationen sortieren nach | | |
| <input type="button" value="übergeordnete Lemmata"/> | <input type="button" value="T scores"/> | | <input type="button" value="T scores"/> |
| Maximalzahl der signifikantesten Kollokationen für die Anzeige im Ergebnis | | | <input type="text" value="10"/> |
| <input type="button" value="Analyse ausführen!"/> | | | |

Abb. 4a: Kollokationsanalyse des Wortes *wmm*, sortiert nach T-scores

Kollokationsanalyse



Kollokationsanalyse für das Lemma *wmm* "essen" (Wb 1, 320.1-321.12) (Lemma-Nummer 46710); die Analyse bezieht sich auf übergeordnete Lemmata; Suchfenster: maximaler Abstand nach links: 0; minimaler Abstand nach links: 0; minimaler Abstand nach rechts: 1; maximaler Abstand nach rechts: 1; Ergebnisausgabe sortiert nach T scores; Ergebnisausgabe beschränkt auf die 10 signifikantesten Kollokationen.

Gesamthäufigkeit des *node* Wortes: 462
Gesamtzahl der Wörter innerhalb des Suchfensters: 454

Liste der Kollokationen, geordnet nach T scores

| Kollokation | gesamte Häufigkeit | Häufigkeit der Kollokation | MI score | T score |
|---|--------------------|----------------------------|----------|---------|
|  =j "[Suffix Pron. sg.1.c.]" (Wb 1, 25; EAG § 159; Schenkel, Einf., 105; ENG §§ 59-64; JWSpG § 216) (Lemma-Nummer 10030) | 17179 | 103 | 3.0445 | 8.9188 |
|  =f "[Suffix Pron. sg.3.m.]" (Wb 1, 572.1; EAG § 159; Schenkel, Einf., 105; ENG § 69; Junge, Näg. Gr., 53) (Lemma-Nummer 10050) | 31068 | 65 | 1.5255 | 5.2618 |
|  =k "[Suffix Pron. sg.2.m.]" (Wb 5, 83.2-3; EAG § 159; Schenkel, Einf., 105; ENG §§ 65-67; Junge, Näg. Gr., 53) (Lemma-Nummer 10110) | 23054 | 49 | 1.5483 | 4.6066 |

Abb. 4b: Ergebnis der Kollokationsanalyse

Bei der Analyse mittels *MI-score* tauchen keine Verbsubjekte auf, sondern nur gegessene Lebensmittel, allerdings oft ziemlich skurrile, die mit Bedrohungen im Jenseits zusammenhängen.

Kollokationsanalyse

node Wort: *wmm* "essen" (Wb 1, 320.1-321.12)

| | | | |
|--|--------------------------------|--------------------------------|---|
| maximaler Abstand nach links | minimaler Abstand nach links | minimaler Abstand nach rechts | maximaler Abstand nach rechts |
| <input type="text" value="0"/> | <input type="text" value="0"/> | <input type="text" value="1"/> | <input type="text" value="1"/> |
| Analyse ausführen für | übergeordnete Lemmata | | Kollokationen sortieren nach |
| | | | MI scores |
| Maximalzahl der signifikantesten Kollokationen für die Anzeige im Ergebnis | | | <input type="text" value="10"/> |
| | | | <input type="button" value="Analyse ausführen!"/> |

Abb. 4c: Kollokationsanalyse des Wortes *wmm*, sortiert nach MI scores

Kollokationsanalyse

 Kollokationsanalyse für das Lemma *wmm* "essen" (Wb 1, 320.1-321.12) (Lemma-Nummer 46710) ; die Analyse bezieht sich auf übergeordnete Lemmata ; Suchfenster: maximaler Abstand nach links: 0; minimaler Abstand nach links: 0; minimaler Abstand nach rechts: 1; maximaler Abstand nach rechts: 1; Ergebnisausgabe sortiert nach MI scores ; Ergebnisausgabe beschränkt auf die 10 signifikantesten Kollokationen.

Gesamthäufigkeit des *node* Wortes: 462
Gesamtzahl der Wörter innerhalb des Suchfensters: 454

Liste der Kollokationen, geordnet nach MI scores

| Kollokation | gesamte Häufigkeit | Häufigkeit der Kollokation | MI score | T score |
|--|--------------------|----------------------------|----------|---------|
|  <i>wzm.w</i> "[Körperteile des Menschen (Innereien?)]" (Wb 1, 357.15) (Lemma-Nummer 49560) | 1 | 1 | 10.4263 | 0.9993 |
|  <i>w4D.y</i> "grüne Pflanzen" (Wb 1, 266.11) (Lemma-Nummer 43600) | 11 | 3 | 8.5518 | 1.7274 |
|  <i>bsk</i> "Eingeweide; Herz" (Wb 1, 477.10-11) (Lemma-Nummer 57520) | 25 | 5 | 8.1044 | 2.2279 |

Abb. 4d: Ergebnis der Kollokationsanalyse

Einstieg über die Texte

Bei der zweiten Einstiegsmöglichkeit in den TLA stehen nicht die einzelnen Lemmata im Vordergrund, sondern der Gesamttext oder eine Gruppe von Texten. Der TLA ermöglicht es, automatisch Indizes zu Texten oder Textgruppen zu erstellen. Der TLA unterscheidet dabei zwischen den Nomina propria und den Nomina Apellativa (Eigennamen und Titel): Bei den Nomina propria kann der Index rein alphabetisch oder nach Wortarten sortiert ausgegeben werden.

Wortindex

Die Analyse bezieht sich auf den folgenden Zweig der Datenbank: **Objekt** pTurin Museo Egizio 1791 Tb 1-113

Analyse ausführen für Ergebnisse anzeigen

alle Belegstellen auflisten

Abb. 5a: Erstellen eines Wortindexes des pTurin Museo Egizio 1791

Wortindex

Die Analyse bezieht sich auf den folgenden Zweig der Datenbank: **Objekt** pTurin Museo Egizio 1791 Tb 1-113
(Das Resultat wurde für übergeordnete Lemmata berechnet.)

[Anfang] [zurück] [weiter] [Ende]

| | | |
|--|--|---|
|  wn "(sich) öffnen" (Wb 1, 311.2-312.11) (Lemma-Nummer 46060) | pTurin Museo Egizio 1791 Tb 1-113, Tb 096 ((Titelzeile)); pTurin Museo Egizio 1791 Tb 1-113, Tb 098 ((6)); pTurin Museo Egizio 1791 Tb 1-113, Tb 099 ((5)); pTurin Museo Egizio 1791 Tb 1-113, Tb 101 ((4)); pTurin Museo Egizio 1791 Tb 1-113, Tb 110 ((11)); pTurin Museo Egizio 1791 Tb 1-113, Tb 110 ((16)); pTurin Museo Egizio 1791 Tb 1-113, Tb 113 ((4)); pTurin Museo Egizio 1791 Tb 1-113, Tb 113 ((5)); | 8 |
|  wn "Fehler; Schuld; Tadel" (Wb 1, 314.7-13; FCD 61) (Lemma-Nummer 46080) | pTurin Museo Egizio 1791 Tb 1-113, Tb 001 ((16)); | 1 |
|  wn "kahl sein" (Wb 1, 314.15-16; FCD 61) (Lemma-Nummer 46100) | pTurin Museo Egizio 1791 Tb 1-113, Tb 084 ((3)); | 1 |
|  wnt "eilen; vorbeigehen; nicht beachten" (Wb 1, 313.10-314.6) (Lemma-Nummer 46280) | pTurin Museo Egizio 1791 Tb 1-113, Tb 011 ((2)); pTurin Museo Egizio 1791 Tb 1-113, Tb 011 ((3)); pTurin Museo Egizio 1791 Tb 1-113, Tb 046 ((2)); | 3 |
|  wnt "Stunde" (Wb 1, 316.1-317.2) (Lemma-Nummer 46420) | pTurin Museo Egizio 1791 Tb 1-113, Tb 005 ((1)); pTurin Museo Egizio 1791 Tb 1-113, Tb 008 ((1)); pTurin Museo Egizio 1791 Tb 1-113, Tb 015 c ((12)); pTurin Museo Egizio 1791 Tb 1-113, Tb 021 ((2)); pTurin Museo Egizio 1791 Tb 1-113, Tb 031 ((2)); pTurin Museo Egizio 1791 Tb 1-113, Tb 032 ((6)); pTurin Museo Egizio 1791 Tb 1-113, Tb 046 ((Titelzeile)); pTurin Museo Egizio 1791 Tb 1-113, Tb 064 ((12)); pTurin Museo Egizio 1791 Tb 1-113, Tb 064 ((12)); | 9 |
|  wnt "sich hin und her bewegen" (Wb 1, 318.1-9) (Lemma-Nummer 46490) | pTurin Museo Egizio 1791 Tb 1-113, Tb 031 ((10)); pTurin Museo Egizio 1791 Tb 1-113, Tb 110 ((13)); | 2 |

Abb. 5b: Ergebnisliste des Wortindexes des pTurin Museo Egizio 1791

Ein solcher Index kann ein Ziel für sich sein, z.B. der „Wortindex zum späten Totenbuch (pTurin 1791)“ von Burkhard Backes⁸, der auf der Grundlage der Digitalisierung dieses Totenbuchs im TLA erstellt wurde. Solche Indizes können aber auch die Ausgangslage für weitere Forschungen sein. Zum Beispiel wurden die im TLA erstellten Indizes von literarischen Texten schon für Clusteranalysen eingesetzt. Simon Schweitzer ist der Frage nachgegangen, ob man auf der Grundlage seines Wortschatzes die Urschrift der „Lehre für König Merikare“, von der nur Abschriften aus dem Neuen Reich überliefert sind, dem literarischen Schaffen des Mittleren oder doch erst des Neuen Reiches zuweisen kann. Nach der hier angewandten Analyse-methode sind die nächstliegenden Texte dieser Lehre zwei Texte – die „Lehre des Ptahhotep“ und der „Beredte Bauer“ –, die unbestritten aus dem Mittleren Reich stammen.⁹

Eine weitere Anwendung, die der TLA bereitstellt, ist die Schlüsselwortanalyse. Diese kann dem Philologen helfen, Thematik und Stil eines Textes genauer zu bestimmen.¹⁰

Eine Schlüsselwortanalyse für die Substantive, die in der Geschichte des „Beredten Bauern“ vorkommen, zeigt folgendes Ergebnis: Es werden Lemmata ausgeworfen, die jeder Ägyptologe bei diesem Text als thematisch wichtig einstufen würde: Der „Bauer“ erleidet „Böses“, weil einer seiner Esel „oberägyptische Gerste“ am Wegrand frisst. Der Getreidebesitzer konfisziert daraufhin unberechtigtweise den Besitz des Bauern. Der Bauer fühlt sich als „Bedrängter“ und versucht in neun Reden beim Distriktverwalter Recht zu bekommen. Diesen vergleicht er u.a. mit einer „Waage“, die im „Gleichgewicht“ sein und nicht absichtlich zu einer Seite abweichen sollte.

⁸ BACKES, B., *Wortindex zum späten Totenbuch (pTurin 1791)*, Studien zum ägyptischen Totenbuch 9, Wiesbaden 2005.

⁹ SCHWEITZER, S., Dating Egyptian literary texts: lexical approaches, in: MOERS, G. *et al.* (eds.), *Dating Egyptian Literary Texts*, Lingua Aegyptiae. Studia Monografica 11, Hamburg 2012.

¹⁰ Schlüsselwörter sind Wörter, die den Text besonders charakterisieren. Sie kommen in einem Text häufiger vor als in anderen Texten und geben so Aufschluss über das Thema oder stilistische Besonderheiten des Textes. Kulturwissenschaftlich gesehen zielt der Begriff auf Wörter, die in kulturellen Diskursen eine zentrale Rolle spielen.

Schlüsselwortanalyse

Die Analyse bezieht sich auf den folgenden Zweig der Datenbank: **Text** Der beredte Bauer (Version B1) pBerlin P 3023 + pAmherst I, Der beredte Bauer (Version B1)

Analyse ausführen für: **übergeordnete Lemmata**

Analyse für die Wortkategorie: **Substantive (ohne Namen und Titel)** Statistiken berechnen mit Bezug auf: **alle Wörter**

minimale Häufigkeit: **5** minimale Abweichung vom Erwartungswert: **2**

Resultat sortieren nach: **Chiquadrat-Statistik** Maximalzahl Wörter im Ergebnis: **25**

Analyse ausführen!

Abb. 6a: Schlüsselwortanalyse für Substantive aus dem beredten Bauern

Schlüsselwortanalyse

Die Analyse bezieht sich auf den folgenden Zweig der Datenbank: **Text** Der beredte Bauer (Version B1) pBerlin P 3023 + pAmherst I, Der beredte Bauer (Version B1) (Das Resultat wurde für übergeordnete Lemmata berechnet.) Analyse für die Wortkategorie: Substantive (ohne Namen und Titel); die Statistiken wurden berechnet mit Bezug auf alle Wörter

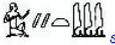
| Lemma | Häufigkeit im Test-Corpus | Häufigkeit im Gesamtkorpus | Abweichung vom Erwartungswert | Chiquadrat Statistik | log-likelihood Statistik |
|---|---------------------------|----------------------------|-------------------------------|-----------------------|--------------------------|
|  <i>sxtj</i> "Feldbewohner, Bauer" (Wb 4, 231.15-232.7; FCD 240) (Lemma-Nummer 141500) | 36 4.6936% | 0.0525% | 35.60 | 3077.88 p=0.000000 | 272.51 p=0.000000 |
|  <i>jyt</i> "das Kommende (euphemist. für Böses)" (Wb 1, 38.9-10) (Lemma-Nummer 21340) | 12 1.5645% | 0.0205% | 11.84 | 823.52 p=0.000000 | 85.65 p=0.000000 |
|  <i>Sma</i> "oberägyptische Gerste („schmale Gerste“)" (Wb 4, 476.8-477.7) (Lemma-Nummer 154800) | 7 0.9126% | 0.0096% | 6.93 | 563.86 p=0.000000 | 53.85 p=0.000000 |

Abb. 6b: Ergebnisliste der Schlüsselwortanalyse für Substantive aus dem beredten Bauern

Eine entsprechende Analyse für die Verben bestätigt diesen Eindruck: Verben des Redens sowie Verben zum Thema „schief sein, betrügen, beseitigen, schädigen“ sind prominent vertreten.

Schlüsselwortanalyse

Die Analyse bezieht sich auf den folgenden Zweig der Datenbank: Text Der beredete Bauer (Version B1) pBerlin P 3023 + pAmherst I, Der beredete Bauer (Version B1)

Analyse ausführen für: übergeordnete Lemmata

Analyse für die Wortkategorie: Verben Statistiken berechnen mit Bezug auf: alle Wörter

minimale Häufigkeit: 5 minimale Abweichung vom Erwartungswert: 2

Resultat sortieren nach: Chiquadrat-Statistik Maximalzahl Wörter im Ergebnis: 25

Abb. 7a: Schlüsselwortanalyse für Verben aus dem beredten Bauern

Schlüsselwortanalyse

Die Analyse bezieht sich auf den folgenden Zweig der Datenbank: Text Der beredete Bauer (Version B1) pBerlin P 3023 + pAmherst I, Der beredete Bauer (Version B1) (Das Resultat wurde für übergeordnete Lemmata berechnet.) Analyse für die Wortkategorie: Verben; die Statistiken wurden berechnet mit Bezug auf alle Wörter

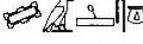
| Lemma | Häufigkeit im Test-Corpus | Häufigkeit im Gesamtcampus | Abweichung vom Erwartungswert | Chiquadrat Statistik | log-likelihood Statistik |
|---|---------------------------|----------------------------|-------------------------------|----------------------|--------------------------|
|  <i>gsA</i> "sich neigen; schief sein" (Wb 5, 205.7-12; FCD 292) (Lemma-Nummer 168510) | 6 0.8299% | 0.0148% | 5.89 | 274.58 p=0.000000 | 39.14 p=0.000000 |
|  <i>awn</i> "betrügen; plündern" (Wb 1, 172.11-18) (Lemma-Nummer 36110) | 7 0.9682% | 0.0360% | 6.74 | 150.83 p=0.000000 | 33.93 p=0.000000 |

Abb. 7b: Ergebnisliste der Schlüsselwortanalyse für Verben aus dem beredten Bauern

Führt man eine solche Schlüsselwortanalyse für alle Lemmata durch, tauchen auch Funktionswörter auf. Diese beleuchten nicht die Thematik, sondern den Stil. Auffällig beim Beredten Bauern ist, dass keine Personalpronomina auftauchen, dafür aber Hervorhebungs-partikel sowie der negative Imperativ „tue nicht!“. Im Gegensatz dazu steht z.B. die autobiographische Erzählung des Sinuhe, in dem die Pronomina „ich“ und „mich“ wichtig sind, sowie indikativische Negativsätze.

Der TLA liefert auch Frequenzdaten von Wörtern und Wortarten in einem Text, wie z.B. Angaben zu den häufigsten Wörtern und

Wortarten oder das *type/token*-Verhältnis.¹¹ Man kann dort die Parameter so hoch stellen, dass nicht nur die häufigsten Wörter, sondern alle Lemmata berücksichtigt werden. Auf diese Weise erhält man ausreichend Daten, um sie für Recherchen im Bereich der Lexikostatistik zu nutzen. Der TLA bietet zum ersten Mal in der Ägyptologie die Möglichkeit, solche Recherchen ohne gigantischen manuellen Zählaufwand durchzuführen.

Die digitalen Rohdaten der altägyptischen Textdatenbank können natürlich auch außerhalb der bisherigen Analysemöglichkeiten des TLA eingesetzt werden. Tatsächlich sind die Möglichkeiten, die in der öffentlich zugänglichen Version der Datenbank geboten werden, für corpuslinguistische Fragestellungen im Augenblick noch relativ beschränkt, ganz im Gegensatz zu dem Potential an Forschungsmöglichkeiten, die die im XML-Format abgelegten Informationen bieten. So hat Katharina Stegbauer (Universität Leipzig) die eingegebenen magischen Texte des Projektes *Digital Heka* mit computerlinguistischen Methoden untersucht, um stilistische Analysen durchzuführen.¹²

Des Weiteren sollen in einem Kooperationsprojekt mit der Ruhr-Universität Bochum und dem Kompetenzzentrum für die Erschließungs- und Publikationsverfahren der Universität Trier ebenfalls die XML-Daten des TLA innerhalb eines Projekts zur Religionssoziologie zur Verfügung gestellt werden. Eine semantische Netzwerkanalyse der in Texten zum Ausdruck gebrachten Relationen zwischen Inhalten und Akteuren soll anhand der digitalen Sprachdaten durchgeführt werden. Im Ergebnis werden so die altägyptischen Relationen mit denen anderer Kulturen auf neuer Basis vergleichbar.¹³

Für die vergleichenden statistischen Analysen eines Textes ist ein Referenzcorpus erforderlich. Im TLA dient die ganze Textdatenbank als Referenzcorpus. Hierbei ist zu bemerken, dass dieses Referenzcorpus bisher kein ausbalanciertes Corpus im Sinne der Corpuslinguistik darstellt.

¹¹ Die *type-token*-Ratio berechnet das Verhältnis zwischen der Zahl der Wörter im Text insgesamt und der Anzahl der einzelnen Lexeme. So sieht man, ob der Text im Verhältnis zur Textlänge einen verhältnismäßig reichen oder armen Wortschatz aufweist. Vgl. SCOTT, M., *Oxford WordSmith Tools*, Oxford 2004.

¹² <http://www.uni-leipzig.de/~digiheka/>.

¹³ <http://aktuell.ruhr-uni-bochum.de/pm2012/pm00248.html.de;>
[http://www.ceres.ruhr-unibochum.de/de/project/forschungsprojekte/senereko/?print=1.](http://www.ceres.ruhr-unibochum.de/de/project/forschungsprojekte/senereko/?print=1)

Die hier beschriebenen Operationen kann man sowohl für das hieroglyphisch-hieratische als auch für das demotische Corpus durchführen. In zukünftigen Projekten ist ein gleichzeitiger Zugriff auf beide Teilcorpora geplant, die dann in einem Gesamtkorpus integriert zur Nutzung aufbereitet werden sollen. In diesem neuen diachronen Ansatz der Sprachforschung soll die wissenschaftshistorisch bedingte Trennung der Sprachstufen des Ägyptischen überwunden werden.¹⁴

¹⁴ Ein entsprechender Antrag für ein Projekt zur diachronen Aufbereitung und Erforschung des Ägyptischen ist bei der Union der Akademien eingereicht und befürwortet worden.