

Simon D. Schweitzer

Eine Hauptkomponentenanalyse mittelägyptischer Literaturwerke

Abstract: In this article, I analyze Middle Egyptian literary texts with a principal component analysis and create six clusters in a cluster analysis: (1) the Hymn to the Nile, (2) the Kemit, (3) the Satire of the Trades, (4) Sinuhe, (5) the Teaching of a Man for his Son, the Loyalist Teaching, the Teaching of Hordjedef and the Sporting King (6) remaining texts. The following words are responsible (more than once!) for the establishment of the clusters: suffix *=fj*, nisba *jmn.j*, particle *jr*, demonstrative pronoun *pw*, preposition *m*, the universal quantifier *nb* and the nouns *nb*, „lord“, *t3*, „land“ and *d3j.w*, „bale of cloth“.

*Zum 65. Geburtstag des geschätzten Stephan J. Seidlmaier,
der seit jeher digitale Innovation sowie gute corpuslinguistische und lexikographische Arbeit zu würdigen weiß.*

1 Einführung

Die wissenschaftlichen Disziplinen produzieren Ergebnisse, Erkenntnisse, Modelle und nicht zuletzt Daten. Es wird prognostiziert, dass im Jahre 2025 das weltweit generierte Datenvolumen 175 Zettabyte umfasst.¹ Natürlich sind in dieser unfassbar großen Zahl nicht nur ägyptologische Forschungsdaten enthalten, aber dennoch sind diese so zahlreich, dass nur eine digitale Auswertung in Frage kommt. Wie man mit archäologischen Forschungsdaten umgehen kann, hat der Jubilar in seiner Dissertation wegweisend demonstriert.² Auch die Erforschung der ägyptischen Sprachdaten³ verdankt ihm viel, und eben solche sollen in diesem Aufsatz in den Blick genommen werden.

Die uns überlieferte, in mittelägyptischer Sprache verfasste Literatur bietet verschiedene Texte recht unterschiedlichen Zuschnitts: Es gibt bspw. fantastische Elemente, Klagen, Rahmenerzählungen. Dementsprechend gibt es viele verschiedene Versuche, die ägyptische Literatur zu gliedern.⁴ Für die Beurteilung solcher Unterteilungen sind m. E. zwei Fragen relevant: 1) Welche Kriterien sind dafür verantwortlich, dass ein Literaturwerk in genau diese Klasse eingesortiert wird? 2) Wie sind die Klassen strukturiert? M. a. W.: Gibt es einen (proto)typischen Vertreter einer Klasse? Sind manche Werke eher in der Peripherie einer solchen Klasse anzusiedeln?

Die Kriterien sind funktionaler Natur, wenn sie bspw. zwischen Bildung und Unterhaltung unterscheiden, rekurren auf Gattungen oder orientieren sich an bestimmten Schlüsselwörtern wie *rl* für lehrhafte Texte. Die Schwierigkeit liegt jetzt nicht in dem Aufbau der Klassen anhand der Kriterien, sondern in der Zuweisung der Literaturwerke zu diesen Klassen. Denn die Zuweisung steht und fällt mit der Interpretation des ägyptischen Textes, die nicht immer eindeutig vorgenommen werden kann. So schreiben Burkard und Thissen in ihrer Einführung über den Schiffbrüchigen: „Die Spanne der Deutungen des *Schiffbrüchigen* unterstreicht die Ratlosigkeit hinsichtlich der Interpretation: Märchen, Seemannsgarn, lehrhafter Text, eschatologische Erzählung, allegorisches Märchen, königliche Propagandaschrift, antikönigliche Oppositionsschrift, Text mit esoterischem Wissen u. a. m. wurden vorgeschlagen.“⁵

Neben diese beschriebene Unsicherheit tritt in fast allen Fällen eine Mischung von Gattungen innerhalb eines Textes: Der Sinuhe ist nicht ausschließlich eine Erzählung, die Klagen des Bauern sind in eine Rahmenerzählung eingebunden. Parkinson schlägt deswegen vor, bei einer Klassifikation von einem „dominant genre“ zu sprechen.⁶

¹ Reinsel et al. 2018.

² Seidlmaier 1990.

³ Der Thesaurus Linguae Aegyptiae, kurz: TLA, (Berlin-Brandenburgische Akademie der Wissenschaften 2004-) ist ohne seine konzeptuellen Arbeiten und seine Programmierung nicht denkbar und würde ohne ihn gar nicht existieren.

⁴ Assmann 1996; Blumenthal 1998; Parkinson 1996, um wenige Beispiele zu nennen.

⁵ Burkard/Thissen 2003, 143.

⁶ Parkinson 1996, 308–309.

Für die Suche nach einem prototypischen Vertreter einer Klasse stellt sich dann die Frage, ob denn der Text mit dem höchsten prozentualen Anteil des dominant genres das zentrale Mitglied dieser Klasse sei.

Im Folgenden soll ein anderes Verfahren vorgeschlagen werden, wie und nach welchen Kriterien ägyptische Literaturwerke gruppiert werden können. Dabei soll auf Verfahren der multivariaten Statistik zurückgegriffen werden, und zwar die Hauptkomponentenanalyse und die darauf basierende Clusteranalyse.

2 PCA

2.1 Einführung

Die Hauptkomponentenanalyse (principal component analysis, im Folgenden kurz: PCA) kann große Datensätze, die viele Beobachtungen in vielen Variablen aufweisen, strukturieren und veranschaulichen.⁷ Dabei werden die vielen Variablen auf die Hauptkomponenten reduziert, die den Datensatz adäquat repräsentieren. Das Prinzip der PCA kann an folgendem Beispiel skizziert werden. Angenommen, man möchte einen Fisch fotografieren. Dann ist eine Seitenansicht einer Vorderansicht vorzuziehen, da die Seitenansicht die Eigenheiten eines Fisches am besten abbildet. Mit einem Foto der Seitenansicht erreicht man das Ziel, den dreidimensionalen Fisch am besten, d. h. mit so wenig Informationsverlust wie möglich, auf einem zweidimensionalen Bild abzubilden. So funktioniert auch die PCA: Die Informationen, die die n Variablen bieten, sollen in wenigen Dimensionen dargestellt werden. Die PCA strukturiert damit den Datensatz in Dimensionen, die sehr inhaltsreich sind, und in Dimensionen, die wenig zum Verständnis der Daten beitragen. Dadurch bietet es sich an, die PCA als Grundlage einer Clusteranalyse zu verwenden. Denn dann kann man sich auf wenige Dimensionen, die das Datenmaterial adäquat abbilden, beschränken, ohne dass die kaum inhaltsreichen höheren Dimensionen die Clusteranalyse evtl. verfälschen.

Wie kann nun die PCA für literarische Werke angewendet werden? Für den Thesaurus Linguae Aegyptiae (TLA) hat die Leipziger Arbeitsstelle des damaligen Akademienvorhabens „Altägyptisches Wörterbuch“ diese ägyptischen Texte bearbeitet, d. h. transkribiert, übersetzt und lemmatisiert. Lemmatisiert heißt, dass jeder Wortform eines Textes das entsprechende Pendant aus der ägyptischen Wortliste zugewiesen wird. Ein n wird also in die Präposition, das Negativwort oder das Genitivwort disambiguierter. Somit kann bei jedem Text bestimmt werden, wie oft die Präposition n vorkommt, wie oft das Negativwort verwendet wird. M. a. W.: es liegen Beobachtungen (=Texte) vor, bei denen bestimmt wird, wie groß eine Variable ist, d. h. wie oft ein Wort aus der ägyptischen Wortliste verwendet wird. Dieser Zusammenhang zwischen Beobachtung (=Text) und Variablen (=Wörter) kann in einer Matrix veranschaulicht werden und bildet so die Voraussetzung, eine PCA durchführen zu können.

2.2 PCA mit 3611 Variablen

Die Daten sind im Jahre 2018 als Datenbankdump im JSON-Format veröffentlicht worden.⁸ Grundlage dieser Untersuchung sind jedoch die daraus konvertierten TEI-Daten,⁹ die als Referenz für das Ancient Egyptian Dictionary¹⁰ publiziert worden sind. Ausgewählt wurden 75 literarische Texte, die in mittelägyptischer Sprache verfasst sind und mehr als 100 Textwörter umfassen. Es ergibt sich ein Datensatz mit 75 Beobachtungen und 3611 Variablen, d. h. die 75 Texte verwenden insgesamt 3611 verschiedene Wörter.¹¹ Vor einer statistischen Anwendung werden die Daten noch normalisiert, da ansonsten die Textlänge Auswirkungen auf die Auswertung haben könnte.¹² Die PCA dieses normali-

⁷ Hier kann aus Platzgründen nicht eine extensive Einführung in die PCA oder in die Clusteranalyse geboten werden. Man vgl. Husson et al. 2017, 1–60 und 173–208.

⁸ Berlin-Brandenburgische Akademie der Wissenschaften 2018.

⁹ Schweitzer 2019a.

¹⁰ Schweitzer 2019b; vgl. Schweitzer 2019c.

¹¹ Der Datensatz ist publiziert unter Schweitzer 2020.

¹² Zu dem positiven Einfluss von (Vektor-)Normalisierungen bei Untersuchungen zu textuellen Daten vgl. Büttner et al. 2017.

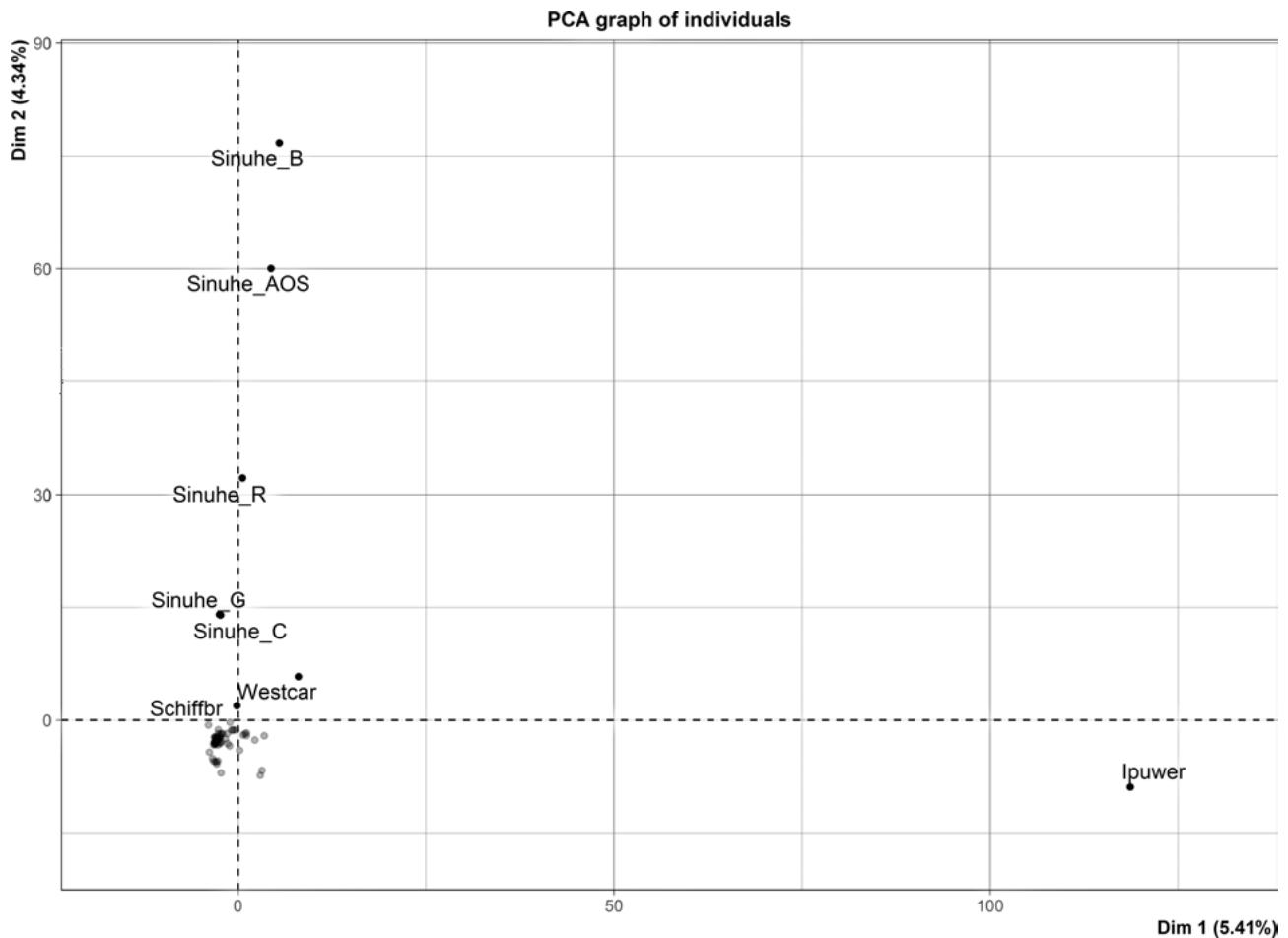


Abb. 1: PCA mit allen Wörtern, erste und zweite Dimension.

sierten Datensatzes wird in der Software R¹³ mithilfe der Pakete FactoMineR¹⁴ und Factoshiny¹⁵ durchgeführt. In Abb. 1 sieht man das graphische Ergebnis der PCA. Die Texte in ihrem 3611-dimensionalen Raum werden auf einer zweidimensionalen Karte dargestellt. Die meisten Texte befinden sich unterhalb der x-Achse in der Nähe des Nullpunktes. Entlang der y-Achse finden sich mit aufsteigenden Werten für die zweite Dimension der Schiffbrüchige, der Papyrus Westcar und fünf Textzeugen der Sinuheerzählung, nämlich C, G, R, AOS und B. Einen hohen Wert in der ersten Dimension und zugleich einen Wert knapp unter Null weist Ipuwer auf. Im Ergebnis liegen also die meisten Texte eng beieinander, während vereinzelte Texte von diesem Ballungszentrum entfernt liegen. In der Tat sind mithilfe des Pakets Factoshiny neun Ausreißer festzustellen, die einen hohen Beitrag zur Konstruktion des Ergebnisses der PCA leisten. Kumuliert man diese Beiträge der Ausreißer, ergibt sich 90,9%, d. h. das Hauptgewicht der Analyse ist durch diese Ausreißer bedingt. Dabei handelt es sich um die Texte Sinuhe B, Ipuwer, Sporting, Westcar, Fishing, Sinuhe AOS, Bauer B1, oDeM 1675 und Bauer R. Es ist beachtlich, dass die neun von der Software bestimmten Ausreißer nicht vollkommen mit den Texten übereinstimmen, die auf Abb. 1 von den anderen Texten entfernt liegen. Diese neun Ausreißer zeichnen sich dadurch aus, dass sie bei bestimmten Variablen signifikant höhere Werte aufweisen als erwartet werden. Die betreffenden Wörter seien im Folgenden summatisch aufgelistet:

Sinuhe B: β , „[Negationspartikel]“ (ID:3), $\beta w.t-c$, „Geschenk; Opfergabe“ (ID:46), βd , „matt werden (des Herzens)“ (ID:351), $=wj$, „[Admirativendung]; [Admirativpartikel]“ (ID:10010), $jmn.tj$, „Westwind“ (ID:26160), $jnr-hd$, „weißer

13 <https://www.r-project.org/> (abgerufen am 24. 06. 2020).

14 FactoMineR <http://factominer.free.fr/> (abgerufen am 24. 06. 2020), vgl. Lê et al. 2008.

15 Factoshiny <http://factominer.free.fr/graphs/factoshiny.html> (abgerufen am 24. 06. 2020).

Stein (Kalkstein; Sandstein)“ (ID:27640), *jšnn*, „Kriegsgeschrei“ (ID:32000), *wčr*, „Flüchtlings“ (ID:44690), *whč*, „fischen und Vögel fangen“ (ID:48800) und *bš.wt*, „Manneskraft“ (ID:53140).

Ipuwer: *šbu*, „brandmarken“ (ID:60), *šfš*, „(gierig) essen“ (ID:119), *štp*, „[ein Kasten]“ (ID:323), *št.t*, „Bahre; Bett“ (ID:334), *šd.w*, „Wut; Aggression“ (ID:346), *šd*, „ausschmieren (eines Topfes mit Ton)“ (ID:349), *jbh.tj*, „Gneis“ (ID:23820), *jph*, „[ein Schwein]“ (ID:24440), *jnd*, „Trauer“ (ID:28060) und *jz.ywt*, „alte Kleider; Lumpen“ (ID:31270).

Sporting: *žh.tj*, „der Horizontbewohner“ (ID:237), *jč*, „Waschnapf; Waschgeschirr“ (ID:21470), *jpn*, „diese [Dem.Pron. pl.m.]“ (ID:24430), *Jr*, „Das Sehen (personifiziert als Gott)“ (ID:28200), *jr.t-šnb.t*, „Brustbehang („was zur Brust gehört“, Schmuck)“ (ID:28450), *jr.w*, „Geschöpf; Gestalt, Gemachtes (als Produkt)“ (ID:29610), *wbš.yt*, „Vorhof (des Tempels)“ (ID:45020), *wnm.j*, „rechts“ (ID:46770), *wnm.w*, „Nahrung“ (ID:46810) und *wntj*, „der Eroberer (meist Apophis, und andere Götter)“ (ID:47150).

Westcar: *Ddj*, „Djedi“ (ID:450047), *jp.t*, „[verschließbarer Gebäudeteil]; [transportabler Schrein]“ (ID:24130), *jstn*, „umschnüren (?)“ (ID:31800), *jdy.t*, „Mädchen“ (ID:33910), *čfn.t*, „Kopftuch („Umhüllung“)“ (ID:37400), *čd.t*, „Kästen; Sargkisten; Kajüte“ (ID:37490), *čmčm*, „(Füße) frottieren“ (ID:37780), *wšg*, „jauchzen (vor Freude); sich freuen“ (ID:43500), *wbš.yt*, „Aufwärterin“ (ID:45030) und *wr-mš.w*, „Größter der Sehenden (Priester)“ (ID:47600).

Fishing: *žjw*, „[atmosphärisches Phänomen (Windstoß?)]“ (ID:9), *žbd.w*, „Abedu (2. Tag nach Neumond)“ (ID:99), *žšr.t*, „Grillklein“ (ID:289), *jn.t*, „Nilbarsch“ (ID:26760), *jtr.tj*, „beide Seiten; (jmds.) Umgebung“ (ID:33320), *č-mh.tj*, „Unterägypten“ (ID:34620), *tp.tj*, „[ein Schiffsteil]“ (ID:40340), *čš-žpd.w*, „steinreich (bildl., „reich an Vögeln“)“ (ID:41070), *wšh*, „Kranz; Girlande (aus Blumen, Gold)“ (ID:43020) und *wjš.t*, „[ein Vogel]“ (ID:44050).

Sinuhe AOS: *žw.t*, „Spende; Darreichung“ (ID:39), *ym*, „Meer“ (ID:24730), *wšj.t*, „breite Halle; Hof; Kapelle“ (ID:49870), *mjn.w.t*, „Hafen“ (ID:68450), *Mnt.w*, „Mentju (Nomadenstämme im NO von Ägypten)“ (ID:71670), *nmw*, „Zwerg“ (ID:84270), *rnn*, „jubeln; preisen“ (ID:95040), *hnt.j*, „vorn; befindlich vor; [lokal]; [temporal]“ (ID:119050), *smsm*, „Ältester“ (ID:135760) und *zhžh*, „laufen; eilen“ (ID:143040).

Bauer B1: *jy*, „[ein Fisch]“ (ID:21320), *jčb*, „(sich) vereinigen; (jmdm. etwas) übergeben; beschenken“ (ID:21680), *jw*, „abtrennen“ (ID:21960), *jšš*, „nehmen; stehlen; (jmdn.) fortführen“ (ID:33530), *č.t*, „das Große; die Größe; das Schwierige“ (ID:34860), *wbb.w*, „[Fische]“ (ID:36100), *čnb.yt*, „Bündel; Korb (Maß für Papyrus und Brot)“ (ID:38400), *čqš.yt*, „Richtigkeit“ (ID:41420), *wčr.w*, „Hast“ (ID:44810) und *wnm.w*, „der Esser; der Verschlinger“ (ID:46820).

oDeM 1675: *žh-bj.t*, „Chemmis; [Name des Pehu-Gebietes im 19. o.äg. Gau]“ (ID:211), *žhžh*, „grün sein; grün machen“ (ID:243), *jfd*, „davonrennen; (durch)eilen“ (ID:24610), *jn.t*, „Tal; Wüstental“ (ID:26780), *jdm.j*, „wertvoller Leinenstoff“ (ID:34030), *ččw*, „aufgeregzt sein; flattern (vom Herzen)“ (ID:35750), *wšd*, „Papyrus (auch als Symbol für U.Äg.)“ (ID:43530), *wždwžd*, „grün sein; grün machen“ (ID:43940), *wnn.yw*, „die Seienden“ (ID:46860) und *bnj*, „Dattel“ (ID:55930).

Bauer R: *žš*, „[eine Nutzpflanze (Wein?)]“ (ID:20250), *jbžš*, „[eine Ölpflanze (ätherisch)]“ (ID:23930), *jnb*, „[eine Pflanze]“ (ID:27200), *jns.t*, „[Pflanze (offizinell)]“ (ID:27830), *čwn.t*, „[ein Baum]; [ein Holz]“ (ID:36140), *čb*, „[ein Mineral]“ (ID:36370), *čn.w*, „Kalksteine (?); Kieselsteine (?)“ (ID:38340), *wbn*, „[eine Pflanze (aus dem Wadi Natrum)]“ (ID:45080), *wnš*, „Schakal“ (ID:47020) und *wgs*, „(ausgenommener?) Vogel“ (ID:50910).

Diese Wörter sind somit insofern charakteristisch für diese Texte, da sie hier prominent oft vertreten sind. Wenn man sie mit den Schlüsselwörtern, wie sie im TLA definiert sind,¹⁶ vergleicht, stellt man fest, dass es praktisch keine Überschneidungen gibt. So sind bspw. die ersten sieben Schlüsselwörter des Sinuhe AOS im TLA: *wh.yt*, „Familie; Stamm“ (ID:48730), *wčr.t*, „Flucht“ (ID:44730), *ms.w-nswt*, „Königskinder (hohe Beamte, kgl. Escorte?); Königsgebäurerinnen (?)“ (ID:75140), *čnb-wšš-snš*, „Leben, Heil, Gesundheit; L.H.G. (Abk.)“ (ID:550035), *=j*, „[Suffix Pron. sg.1.c.]“ (ID:10030), *č-hn.wtj*, „Kabinett; Audienzhalle“ (ID:850364) und *Rtn.w*, „Retjenu (Syrien-Palästina)“ (ID:96590). Die in der PCA ermittelten Wörter, die eine im Vergleich zu den anderen Texten höhere Häufigkeit aufweisen, sind jedoch hauptsächlich seltene Wörter, die zum Teil nur in betreffendem Text überliefert sind oder die innerhalb der literarischen Texte nur dort vorhanden sind. Augenscheinlich werden bei einer PCA, die alle Wörter eines

¹⁶ Seidlmayer/Hafemann 2011, 47: „Statistisch gesehen sind Schlüsselwörter Wörter, die in einem Text oder Teilcorpus häufiger als in anderen Textcorpora oder im Sprachgebrauch insgesamt vorkommen [...].“

Textes heranzieht, die seltenen Wörter, die nur in einem Text überliefert sind, in einem zu starken Maße berücksichtigt, sodass in diesem Fall ein solch heterogenes Ergebnis erzielt wird, in dem mehr als 10 %, nämlich neun Texte, von den insgesamt 75 Texten als Ausreißer zu klassifizieren sind.

2.3 PCA mit 715 Variablen

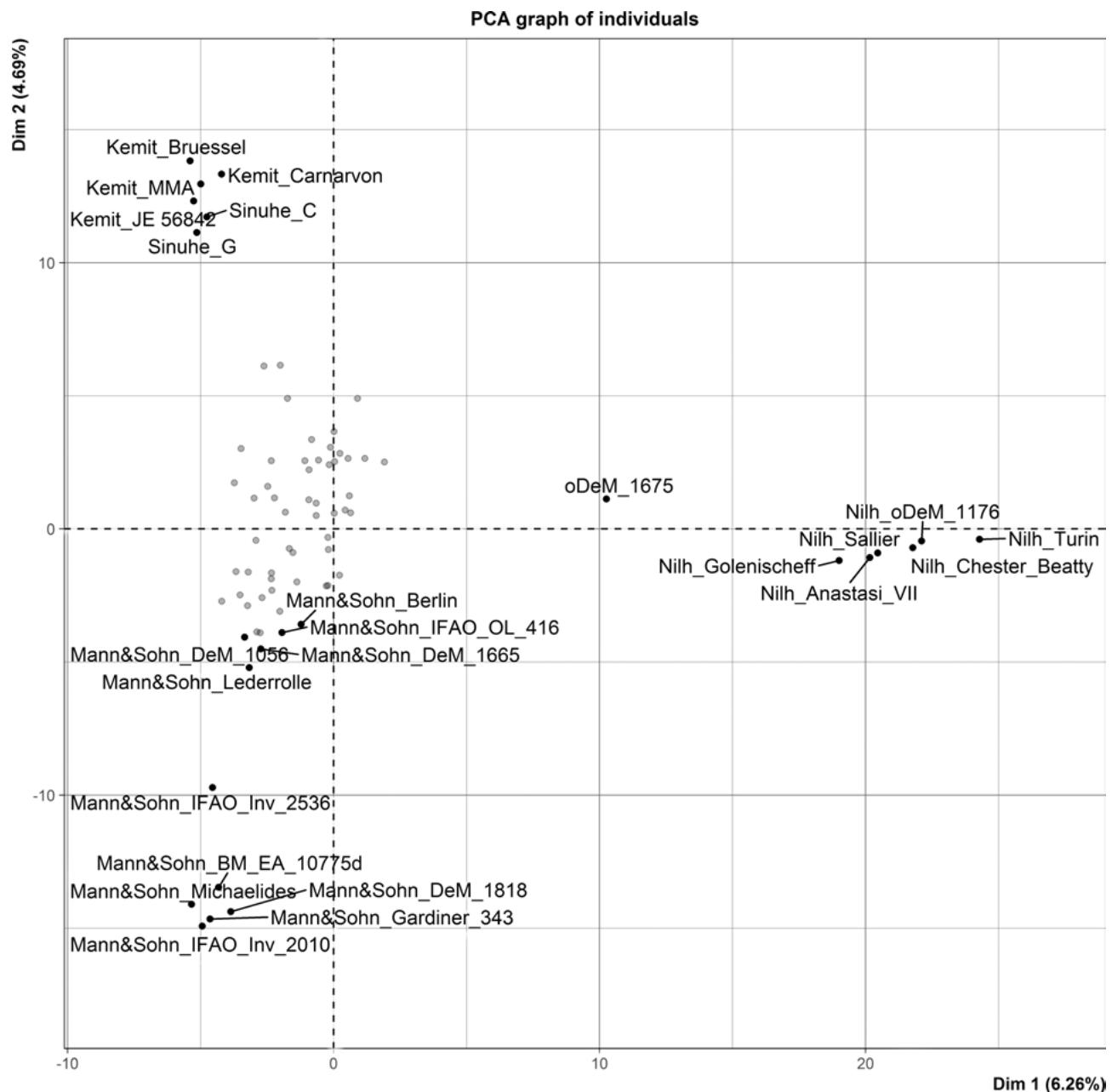


Abb. 2: modifizierte PCA ohne die seltenen Wörter, erste und zweite Dimension.

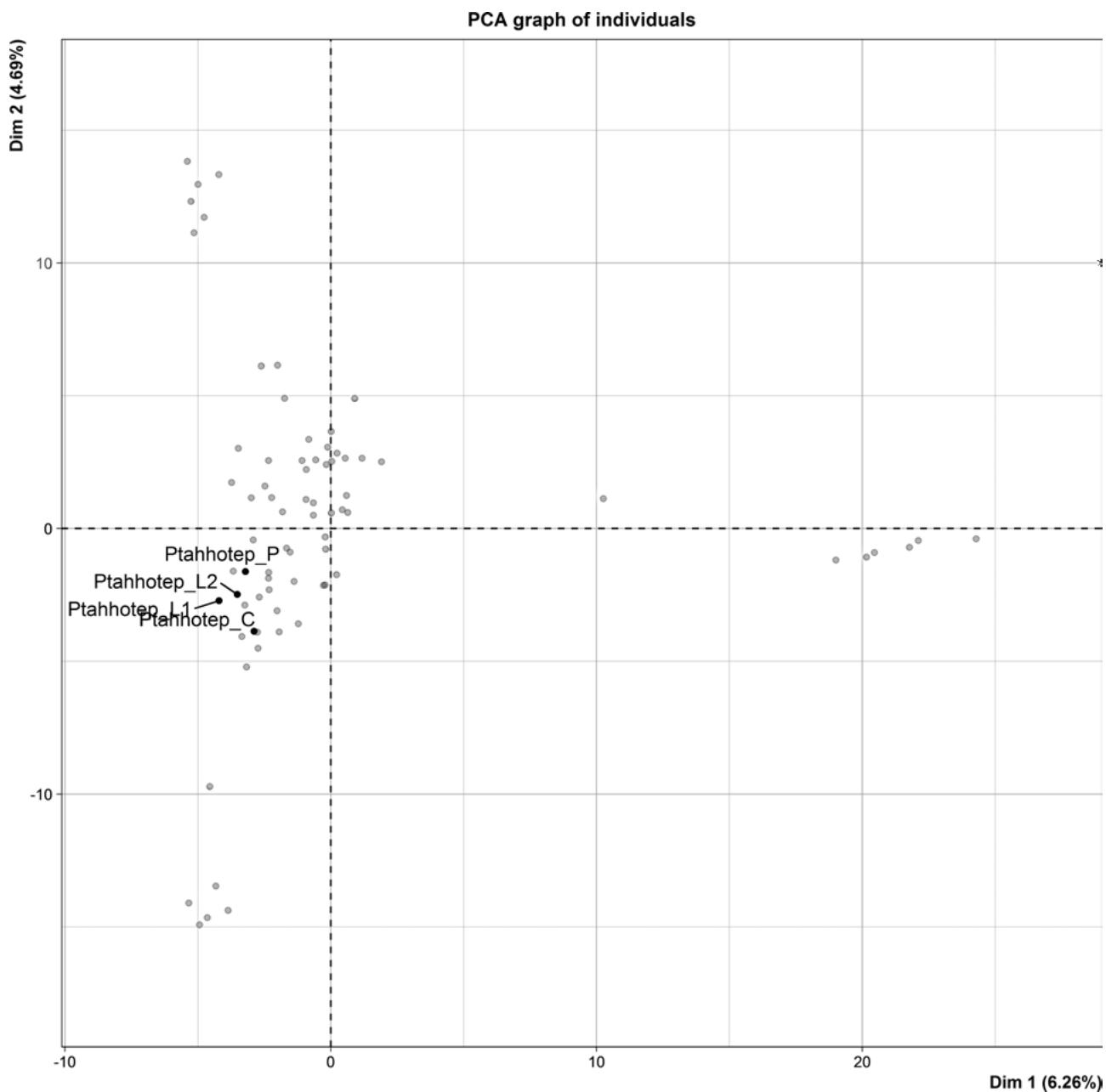


Abb. 3: modifizierte PCA ohne die seltenen Wörter, erste und zweite Dimension, Textzeugen der Lehre des Ptahhotep.

Stilometrische Untersuchungen, die sich das Ziel der Autorenklassifikation gesetzt haben, verzichten oft auf die seltenen Wörter. Das in diesem Bereich vor allem eingesetzte Burrows Delta beschreibt den stilistischen „Fingerabdruck“ eines Autors mithilfe der 100–5000 häufigsten Wörter.¹⁷ Es bietet sich also an, auch den Datensatz der literarischen Texte zu modifizieren, indem die seltenen Wörter ausgeblendet werden, sodass die oben beschriebene Inhomogenität durch das Übergewicht der seltenen Wörter nicht entsteht. Dazu wurden für den modifizierten Datensatz nur Wörter berücksichtigt, die in dem Corpus der 75 literarischen Texte mindestens 10 Belege aufweisen. Die Anzahl der Dimensionen reduziert sich dadurch erheblich: Statt durch 3611 Variablen werden die Texte nun durch 715 beschrieben, was eine Reduktion auf rund ein Fünftel bedeutet. Auch der modifizierte Datensatz wird

¹⁷ Burrows 2002.

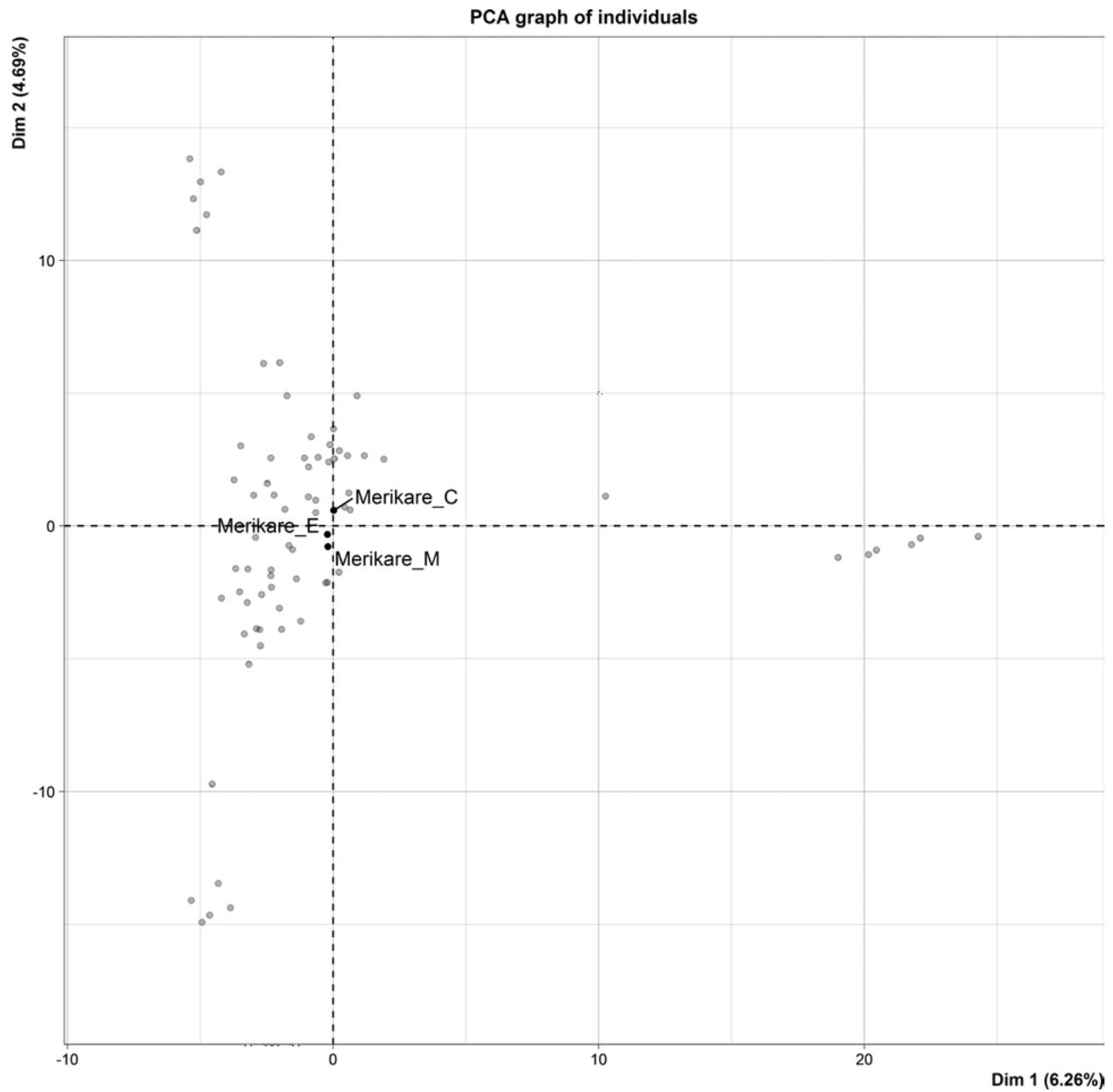


Abb. 4: modifizierte PCA ohne die seltenen Wörter, erste und zweite Dimension, Textzeugen der Lehre für Merikare.

normalisiert. Das Ergebnis der erneuten PCA ist in Abb. 2 zu sehen. Dass das Ergebnis nicht unplausibel ist, lässt sich dadurch verdeutlichen, wenn man sich die Verteilung von Textzeugen eines Werkes anschaut. In Abb. 3 und 4 sieht man die Textzeugen des Ptahhotep bzw. des Merikare. Lägen diese weit entfernt, spräche das gegen den Wert der PCA. Die zu erwartende Verbundenheit drückt sich auch durch die Nähe der Punkte aus.

Somit liefert die PCA Ergebnisse, die zu genauerer Analyse einladen: Man erkennt vier voneinander abgrenzbare Bereiche: Die Textzeugen des großen Nilhymnus weisen hohe Werte in der ersten Dimension und leicht negative in der zweiten auf. Die Textzeugen der Kemit und zwei Textzeugen des Sinuhe besitzen leicht negative Werte in der ersten Dimension und hohe in der zweiten. Die Textzeugen der Lehre eines Mannes für seinen Sohn weisen allesamt leicht negative Werte in der ersten Dimension und negative Werte in der zweiten Dimension auf, wobei einige stark negative Werte in der zweiten Dimension haben. Alle übrigen Texte sind um den Nullpunkt gruppiert, bis auf oDeM 1675, der einen Nilhymnus überliefert. Dieser Text liegt zwischen der Gruppe um den Nullpunkt und der

Gruppe der Textzeugen des großen Nilhymnus. Die unterschiedlichen Gruppen zeichnen sich durch die unterschiedliche Verwendung einzelner Wörter aus. Die Textzeugen des großen Nilhymnus sind charakterisiert durch die signifikant höhere Verwendung der Wörter *hrp*, „leiten; beaufsichtigen“ (ID:120150), *h̄s.t.*, „Bergland; Fremdland; Wüste“ (ID:114300), *H̄pj*, „Hapi“ (ID:650066), *sw̄d*, „grünen lassen; gedeihen lassen“ (ID:129930), *rmt*, „Mensch; Mann“ (ID:94530), *wbn*, „aufgehen; überquellen“ (ID:854500), *sm.w*, „Kraut; Futterkraut; Gemüse“ (ID:134140), *nb*, „jeder; alle; irgendein“ (ID:81660), *dw̄*, „früh auf sein; preisen; anbeten“ (ID:854584) und *jt*, „Gerste; Korn (allg.)“ (ID:32830). Man ist nicht überrascht, diese Wörter als typische Vertreter des großen Nilhymnus anzutreffen. Vergleicht man aber diese charakteristischen Wörter mit den Schlüsselwörtern, die der TLA für den großen Nilhymnus anbietet, so findet man nur die Übereinstimmung mit dem ersten Schlüsselwort *H̄pj*, „Hapi“ (ID:650066). Die nächsten sechs Schlüsselwörter *mnmn.t*, „Herde; Vieh“ (ID:70730), *k̄p*, „bedecken; überdachen; (sich) verbergen“ (ID:854567), *zb̄t*, „Gelächter“ (ID:131660), *wp.wt*, „Schar von Frauen (zur Begrüßung u. Ä.)“ (ID:45740), *bk̄t*, „Schwangere“ (ID:57840) und *nn*, „[Negationspartikel]“ (ID:851961) fehlen in den in der PCA als signifikant häufig klassifizierten Wörtern. Dieser Unterschied lässt sich einerseits gut durch die unterschiedlichen Vergleichstexte erklären. Im TLA werden alle Texte herangezogen, um als Kontrastfolie zu dienen, die spezifischen Wörter des großen Nilhymnus zu ermitteln. In der PCA stehen aber nur 75 literarische Texte zur Verfügung. Es ist zu erwarten, dass unterschiedliche Referenzcorpora unterschiedliche Ergebnisse nach sich ziehen. Das lässt sich gut an *sw̄d*, „grünen lassen; gedeihen lassen“ (ID:129930) verdeutlichen, das die PCA als charakteristisch für den großen Nilhymnus bestimmt. Dieses Wort ist im Corpus der 75 literarischen Texte nur in den Textzeugen des großen Nilhymnus belegt. Andererseits sind die in der PCA gewonnenen Wörter des Nilhymnus charakteristisch für die Positionierung dieser Gruppe innerhalb der ersten beiden Dimensionen, die natürlich nur einen Teil der Information abbilden können.

Die PCA ermittelt nicht nur, welche Wörter überdurchschnittlich in einer Gruppe verwendet werden. Ebenso kann angegeben werden, welche Wörter seltener als erwartet in der Gruppe erscheinen. Für die Textzeugen des großen Nilhymnus sind es folgende Wörter: *jb*, „Herz; Verstand; Charakter; Wunsch“ (ID:23290), *jw*, „[aux.]“ (ID:21881), *dd*, „sagen; mitteilen“ (ID:185810), *hrw*, „Stimme; Geräusch; Zank; Krach“ (ID:120010), *ʒj*, „herrlich; nützlich; verklärt“ (ID:600475), *hpr (m-s)*, „jmd. verfolgen“ (ID:858535), *gr.w*, „Schweigsamer“ (ID:167800), *zb̄j*, „(weg)gehen; führen; aussenden; durchlaufen; verbrennen“ (ID:131460), *ɔɔ*, „viel sein; zahlreich sein; reich sein“ (ID:41010) und *sw̄j*, „vorbeigehen; passieren“ (ID:129740). Dadurch wird zum einen der formale Aufbau des Textes beschrieben, denn Redeeinleitungen mit *dd* oder Sätze mit *jw* sind selten. Zum anderen erkennt man inhaltliche Aspekte, beispielsweise dass der große Nilhymnus nicht zum weisheitlichen Diskurs zu zählen ist, da der *gr.w* geringe Werte aufweist, oder dass die Bewegung von Protagonisten eine geringe Rolle spielt, da *zb̄j* und *sw̄j* unterdurchschnittlich selten sind.

Die Kemitgruppe zeichnet sich durch die hohe Verwendung folgender Wörter aus: *wpw.tj*, „Bote“ (ID:45760), *=j*, „[Suffix Pron. sg.1.c.]“ (ID:10030), *n.j*, „gehörig zu“ (ID:850787), *z̄s*, „ausbreiten; ausstreuern“ (ID:144320), *n-zp*, „nie-mals; [Negationswort]“ (ID:79450), *jr*, „[Partikel (nachgestellt zur Betonung)]“ (ID:28170), *pn*, „dieser [Dem.Pron. sg.m.]“ (ID:59920), *jm*, „da; dort; davon; dadurch; damit“ (ID:24640), *hrw*, „Tag“ (ID:99060) und *jnk*, „ich [Selbst. Pron. sg.1.c]“ (ID:27940). Bemerkenswert ist hier die hohe Anzahl von semantisch schwachen Funktionswörtern. Signifikant niedrige Werte haben in der Kemitgruppe folgende Wörter: *=tw*, „[Suffix Pron. sg.3.c.]“ (ID:170100), *h̄j*, „herabsteigen; fallen; zu Fall kommen“ (ID:97350), *nn*, „[Negationspartikel]“ (ID:851961), *ɔ*, „Erde; Land (als Element des Kosmos); Land (geogr.-polit.); Ägypten; Erdreich (stofflich); Ackerboden; bebaubares Land; [ein Flächenmaß]“ (ID:854573), *mh*, „voll sein; füllen; fassen; packen; auslegen (mit Steinen)“ (ID:854514), *pri*, „herauskommen; herausgehen“ (ID:60920), *wzf*, „träge sein; (etwas) vernachlässigen“ (ID:49520), *tm*, „zu Ende sein“ (ID:854578), *n.tj*, „der welcher (Relativpronomen)“ (ID:89850) und *ʒbj*, „wünschen; sich freuen“ (ID:73).

Die Gruppe der Textzeugen der Lehre eines Mannes für seinen Sohn zeichnet sich durch die starke Verwendung folgender Wörter aus: *ns*, „Zunge“ (ID:87800), *r'*, „Mund“ (ID:92560), *pw*, „[Dem Pron. sg.m.]“ (ID:851517), *hpr (m-s)*, „jmd. verfolgen“ (ID:858535), *m*, „nicht (Imperativ des Negativverbs *jm̄j*)“ (ID:64410), *gr.w*, „Schweigsamer“ (ID:167800), *mrw.t*, „Liebe; Beliebtheit; Wunsch“ (ID:72650), *hn*, „Rede; Spruch; Angelegenheit“ (ID:117520), *wd̄c*, „(ab)trennen; richten; zuweisen“ (ID:52360) und *qbb*, „kühl sein; kühnen“ (ID:160170). Hier findet man Vokabular des weisheitlichen Diskurses. Bemerkenswert ist ferner, hier das Demonstrativum *pw* zu finden, während die Kemitgruppe, die in der zweiten Dimension die entgegengesetzten Werte im Vergleich zu der Gruppe um die Lehre eines Mannes für seinen Sohn aufweist, durch das Demonstrativum *pn* gekennzeichnet ist. Signifikant niedrige Werte haben in dieser Gruppe die folgenden Wörter: *rd̄j*, „geben; veranlassen“ (ID:851711), *ʒpd*, „Vogel (allg.); Geflügel (koll.)“ (ID:107), *jr̄j*, „machen“ (ID:851809), *hrw*, „Tag“ (ID:99060), *jb.t*, „Sache; Besitz; Speisen; Opfer; Reliquie“

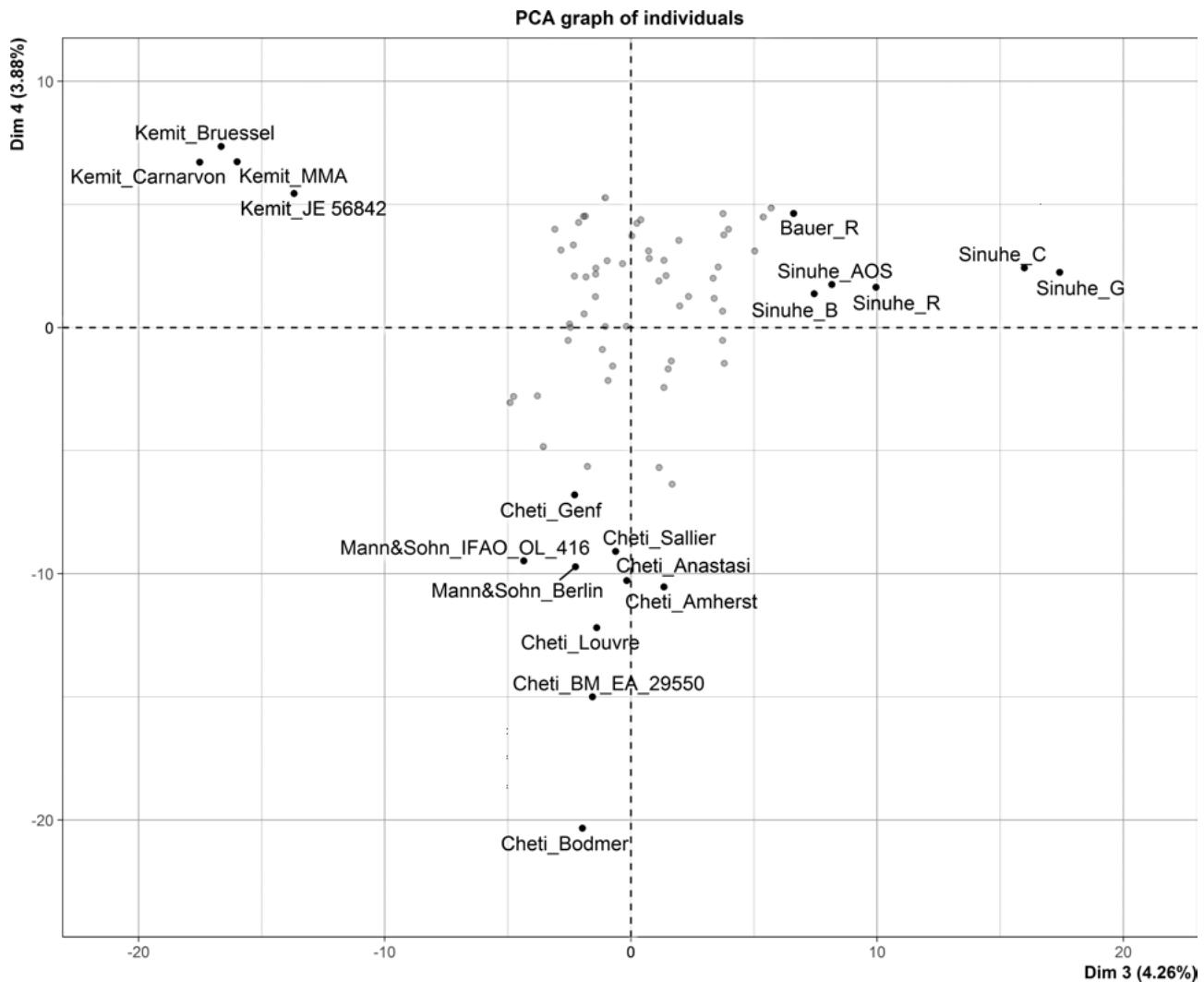


Abb. 5: modifizierte PCA ohne die seltenen Wörter, dritte und vierte Dimension.

(ID:30750), *m*, „[Präposition]“ (ID:64360), *ntr*, „Gott“ (ID:90260), *Ptḥ*, „Ptah“ (ID:62980), *nmḥ.w*, „Armer; Bürger; Waise“ (ID:84370) und *wr*, „groß; viel; reich; bedeutend“ (ID:47271). Charakteristisch ist also das geringe Vorkommen einiger hoch frequenter Wörter wie die Präposition *m*, die Verben *rđi*, *jri*, und das Substantiv *ntr*.

Abb. 5 gibt die Dimensionen 3 und 4 wieder. Hier erkennt man die Streuung der Texte in drei Richtungen. Sinuhetextzeugen haben hohe Werte in der dritten Dimension, die Kemitgruppe hat niedrige Werte. Beide Gruppen haben positive Werte in der vierten Dimension. Stark negative Werte in der vierten Dimension weist die Chetigruppe auf. Typisch für die Sinuhetextzeugen sind die hohen Werte für: *ph*, „erreichen; angreifen“ (ID:61370), *z*, „Mann; Person; Wesen“ (ID:125010), *sdm*, „hören“ (ID:150560), *w^c*, „einer (von mehreren); etwas (von e. Material)“ (ID:600041), *gs*, „Seite; Hälfte“ (ID:854572), *jti*, „nehmen; ergreifen; erobern“ (ID:33560), *t*, „Erde; Land (als Element des Kosmos); Land (geogr.-polit.); Ägypten; Erdreich (stofflich); Ackerboden; bebaubares Land; [ein Flächenmaß]“ (ID:854573), *wj*, „ich; mich [Enkl. Pron. sg.1.c]“ (ID:44000), *nswt-bj.tj*, „König von Ober- und Unterägypten; König von OÄg. u. UÄg. (Thronname der Königstitulatur)“ (ID:88060) und *chc.n*, „[aux.]“ (ID:40111). Diese Wörter spiegeln gut die erzählerischen Elemente der Geschichte wieder: die Strukturierung durch *chc.n*, die Reise mit *ph* oder *t* und die Ich-Erzählung mit *wj*. Unterrepräsentiert sind die Wörter: *dj.w*, „Leinenstoffballen; Stoffballen“ (ID:177680), *=fj*, „[Suffix Pron. sg.3.m. nach Dual]“ (ID:10060), *wd*, „unversehrt sein“ (ID:52090), *hr*, „[Präposition]“ (ID:107520), *rm*, „Fisch (allg.)“ (ID:94160), *m^s.t*, „Knie; Haxe (beim Tier)“ (ID:67370), *sndm*, „angenehm machen; ruhen“ (ID:851678), *sp.t*, „Lippe; Rand; Ufer“ (ID:132440), *z_b.w*, „Schreiber“ (ID:550055) und *hr*, „also; aber; denn; und danach (Konjunktion); [Partikel]“ (ID:119600).

Die Kemitgruppe war auch in den ersten beiden Dimensionen als eigenständige Gruppe erkennbar. Verantwortlich für die Selbstständigkeit in der dritten und vierten Dimension ist die hohe Anzahl folgender Wörter: *nn-n*, „diese [Dem. Pron. pl. c.]“ (ID:500001), *hzi*, „loben; begünstigen“ (ID:109620), *hnt.j*, „Vorhalle“ (ID:119070), *wrh*, „aufstreichen; salben“ (ID:48030), *nt.w*, „Myrrhenharz (für Salböl und Weihrauch)“ (ID:39010), *zj*, „gehen“ (ID:127740), *mir*, „lieben; wünschen“ (ID:72470), *hm.t*, „Frau; Ehefrau“ (ID:104730), *grh*, „Nacht“ (ID:167920) und *k*, „Ka; Lebenskraft“ (ID:162870). So wird sehr gut erkennbar, dass die einzelnen Wörter als Variablen in unterschiedlichem Maße die Dimensionen beeinflussen; denn es finden sich in den Wörtern, die in den ersten beiden Dimensionen typisch für die Kemit sind und diese als eigene Gruppe von den übrigen Texten separieren, keine Übereinstimmungen zu den Wörtern aus der dritten und vierten Dimension. Insofern darf man diese Wörter auch nicht als Schlüsselwörter für die Textgruppe interpretieren; denn ein Wort könnte ein Schlüsselwort eines Textes sein, aber keinerlei bis kaum Einfluss auf die Konstruktion der ersten Dimensionen haben, sodass dieses Wort nicht berücksichtigt wird. Charakteristisch für die Kemitgruppe in der dritten und vierten Dimension ist der niedrige Wert der Anzahl folgender Wörter: *nn*, „[Negationspartikel]“ (ID:851961), *=f*, „[Suffix Pron. sg.3.m.]“ (ID:10050), *rh*, „wissen; kennen; erkennen; wissen dass (mit Verbform)“ (ID:95620), *=tw*, „[Suffix Pron. sg.3.c.]“ (ID:170100), *r*, „[Präposition]“ (ID:91900), *pw*, „[Dem Pron. sg.m.]“ (ID:851517), *hpr (m-s)*, „jmd. verfolgen“ (ID:858535), *mw*, „Wasser“ (ID:69000), *m*, „[Präposition]“ (ID:64360) und *b*, „Erde; Land (als Element des Kosmos); Land (geogr.-polit.); Ägypten; Erdreich (stofflich); Ackerboden; bebaubares Land; [ein Flächenmaß]“ (ID:854573). Die geringen Werte von *nn*, *=tw* und *b* sind auch in den ersten beiden Dimensionen für die Kemitgruppe charakteristisch. Hinzu kommt *rh*, womit sich die Kemit vom weisheitlichen Diskurs separiert.

Die Chetigruppe zeichnet sich durch die hohe Verwendung der folgenden Wörter aus: *=fj*, „[Suffix Pron. sg.3.m. nach Dual]“ (ID:10060), *=f*, „[Suffix Pron. sg.3.m.]“ (ID:10050), *sw*, „er [Enkl. Pron. sg.3.m.]; sich [Enkl. Pron. sg.3.m.]; ihn [Enkl. Pron. sg.3.m.]“ (ID:129490), *hr*, „[Präposition]“ (ID:107520), *hr*, „unter; infolge von“ (ID:850794), *jw.t*, „Funktion; Amt; Würde“ (ID:20430), *pr*, „Haus; Palast; Tempel; Grab; Verwaltung“ (ID:60220), *db*, „Finger“ (ID:183430), *mw*, „Wasser“ (ID:69000) und *shn*, „mischen; sich mischen (unter)“ (ID:153490). In geringem Maße sind die folgenden Wörter in der Chetigruppe vorhanden: *n.j*, „gehörig zu“ (ID:850787), *=j*, „[Suffix Pron. sg.1.c.]“ (ID:10030), *jb*, „Herz; Verstand; Charakter; Wunsch“ (ID:23290), *pn*, „dieser [Dem.Pron. sg.m.]“ (ID:59920), *zbi*, „(weg)gehen; führen; aussenden; durchlaufen; verbrennen“ (ID:131460), *tw*, „du; dich [Enkl. Pron. sg.2.m.]“ (ID:174900), *wj*, „ich; mich [Enkl. Pron. sg.1.c.]“ (ID:44000), *ji*, „machen“ (ID:851809), *wpw.tj*, „Bote“ (ID:45760) und *ky*, „anderer“ (ID:163760). Auffällig ist vor allem die Verwendung der Personen in der Lehre des Cheti. Die dritte Person wird besonders in den Vordergrund gerückt: *=fj*, *=f* und *sw*, während die erste und die zweite Person in Form der Wörter *=j*, *tw* und *wj* stark zurücktritt.

Bei der Charakterisierung der Gruppen ist bemerkenswert, dass einige Wörter mehr als einmal zur Gruppenbildung beitragen. Folgende Tabelle stellt diese Wörter und ihren Einfluss auf die Texte zusammen, wobei *+* bedeutet, dass das betreffende Wort signifikant häufig verwendet wird, und *-*, dass es signifikant selten verwendet wird:

Tab. 1: Charakteristische Wörter und ihr Einfluss auf die Gruppenbildung.

Nilhymnus	Kemit	Lehre eines Mannes	Sinuhe	Cheti
<i>=j</i>	+			-
<i>=f</i>	-			+
<i>=fj</i>		-		+
<i>jb</i>	-			-
<i>jri</i>		-		-
<i>wj</i>		+		-
<i>wpw.tj</i>	+			-
<i>pw</i>	-	+		
<i>pn</i>	+			-
<i>m</i>	-	-		
<i>mw</i>	-			+
<i>n.j</i>	+			-
<i>hrw</i>	+	-		
<i>hr</i>			-	+
<i>hpr</i>	-	-	+	

Tab. 1 (fortgesetzt)

	Nilhymnus	Kemit	Lehre eines Mannes	Sinuhe	Cheti
zbi	–			–	
gr.w	–		+		
β		–		+	

Die PCA ermöglicht es nicht nur, den Informationsgehalt von vielen auf wenige Dimensionen herunterzubrechen, sondern auch die Variablen zu bestimmen, die vorrangig zur Beschreibung der Differenzierung der Literaturwerke beitragen, weswegen sie ja auch den Namen *Hauptkomponentenanalyse* trägt. Dies sind die in dieser Tabelle aufgeführten Wörter, die sich somit als Beschreibungsinstrumentarium für literarische Texte eignen. Ob ein Literaturwerk eines dieser Wörter signifikant häufig oder signifikant selten verwendet, ist ein guter Indikator der Nähe des Werkes zu den oben beschriebenen Gruppen, wenn diese Eigenheiten geteilt werden. Oder anders gesagt: welche Demonstrativreihe in einem Text verwendet wird oder inwieweit indirekte Genitivkonstruktionen in Gebrauch sind, ist für eine Klassifikation eines Literaturwerkes wirklich bedeutsam. Die Benutzung von den Präpositionen *m* und *hr* ist für eine Beurteilung relevant, die Benutzung von *r* dagegen nicht.

3 Clusteranalyse

3.1 Dendrogramm der Clusteranalyse

Die Ergebnisse der PCA können nun für eine Clusteranalyse weiterverwendet werden. Wie auf Abb. 6 deutlich zu sehen ist, kann man auch hier die fünf aus der PCA bekannten Gruppen Nilhymnus, Kemit, Lehre eines Mannes, Sinuhe und Cheti wiederfinden, sodass man insgesamt sechs Cluster erkennen kann. Die Anzahl, wie viele Cluster gebildet werden, ist nicht vorgegeben. Wenn man die Zahl auf sieben, acht oder neun Cluster erhöht, wird aber interessanterweise nicht der Cluster der übrigen Texte aufgespalten. Stattdessen werden die Cluster Lehre eines Mannes und Sinuhe aufgetrennt. Die Cluster selbst sind nicht gleich verteilt: Der Nilhymnuscluster besteht aus sieben Textzeugen, der Kemitcluster aus vier, der Cluster mit Vertretern der Lehre eines Mannes für seinen Sohn aus 20, der Sinuhecluster aus fünf, der Cheticcluster aus zehn und der Cluster mit den restlichen Texten aus 29. Als gut zu separierende Cluster kann man gemäß der Texte, die sich darin befinden, den Nilhymnuscluster, den Kemitcluster und den Cheticcluster ansprechen. Zum einen befinden sich die Textzeugen dieser Texte ausschließlich in einem gemeinsamen Cluster, und zum anderen sind in diesen Clustern keine anderen Texte anzutreffen mit Ausnahme von oDeM 1675, der zum Nilhymnuscluster gehört, was nicht verwundern kann, da es sich bei diesem Text um den einzigen weiteren Nilhymnus neben dem großen Nilhymnus handelt. Der Sinuhecluster besteht auch ausschließlich aus Textzeugen der Sinuhezählung, aber Sinuhe_L ist dort nicht zu finden. Auch eine frühere Clusteranalyse der literarischen Texte brachte die Sonderstellung dieses Textzeugen zu Tage.¹⁸ Sinuhe_L befindet sich stattdessen im Cluster, in dem sich alle Vertreter der Lehre eines Mannes für seinen Sohn, alle Vertreter der Loyalistischen Lehre, die Lehre des Hordjedef und die Erzählung Sporting King befinden. Da es sich bei den Texten, die dem Zentrum des Clusters am nächsten sind, um Zeugen der Lehre eines Mannes für seinen Sohn handelt, sei dieser Cluster nach diesem Text und nicht nach der Loyalistischen Lehre benannt, obgleich beide Texte all ihre Textzeugen in diesem Cluster versammeln. Der größte Cluster umfasst auch die größte Vielfalt der Texte: Lehren wie Ptahhotep oder Merikare, Klagen wie Chacheperreseneb oder Ipuwer, Erzählungen wie der Schiffbrüchige oder Westcar.

Auch wenn diese Aufteilung auf den ersten Blick überrascht, ist sie keineswegs unplausibel. Das wäre sie, wenn sich Vertreter eines Textes auf mehrere Cluster verteilen. Die Plausibilität erhöht sich noch weiter, wenn man diese Clusteranalyse mit einer schon publizierten vergleicht, die ebenfalls die literarischen Texte zur Grundlage hatte.¹⁹

¹⁸ Schweitzer 2013, 185, Fig. 3.

¹⁹ Schweitzer 2013.

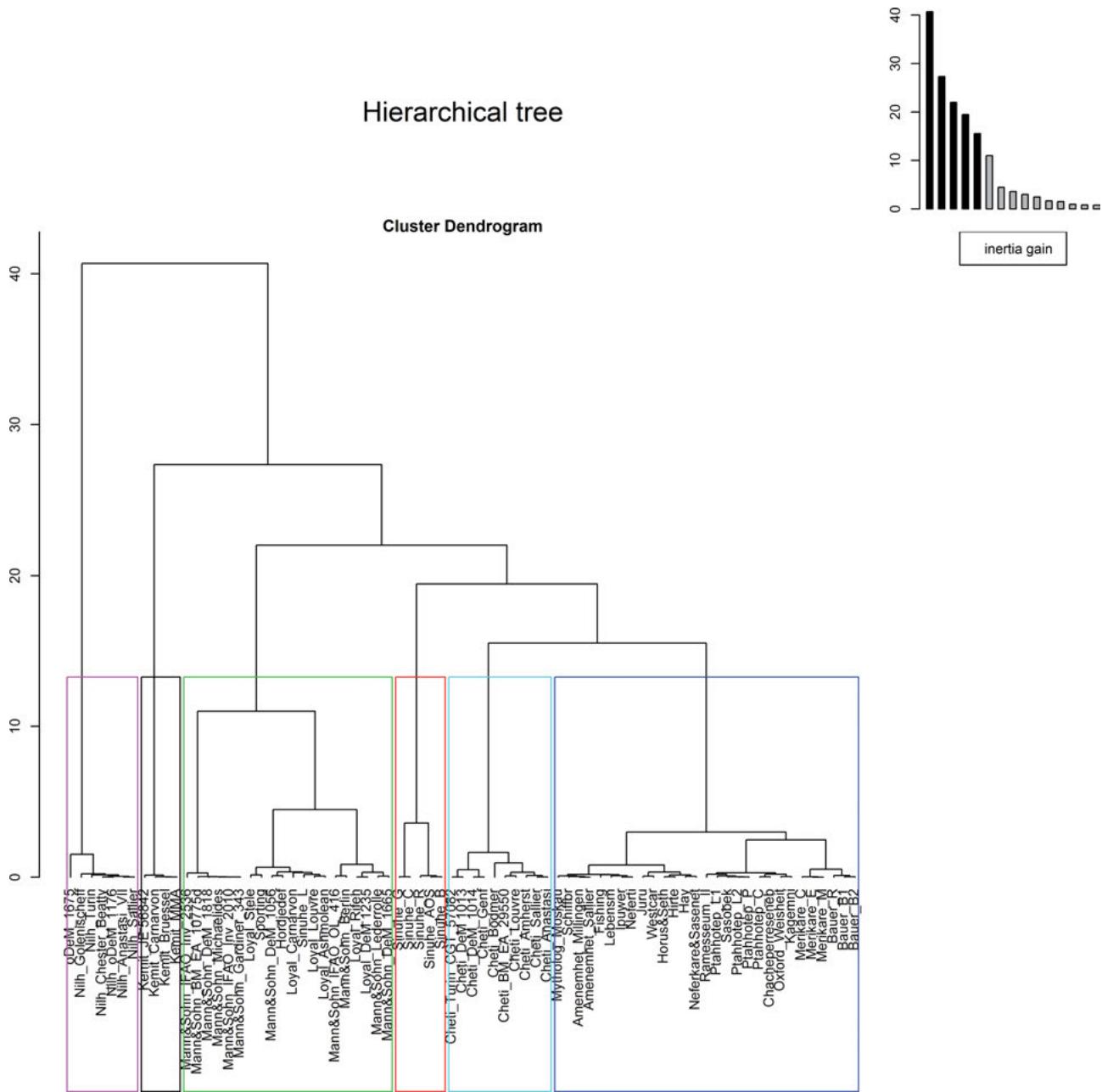


Abb. 6: Dendrogramm der Clusteranalyse.

Die Ergebnisse differieren, was aber aus verschiedenen Gründen erklärlich ist. Zunächst ist die Datenbasis unterschiedlich. Zwar verwenden beide Analysen die mittelägyptisch geschriebenen Literaturwerke ab einer bestimmten überlieferten Textlänge, aber die Daten, die in dieser Untersuchung benutzt werden, sind im Rahmen der Erstellung des Ancient Egyptian Dictionary bereinigt worden. Dies umfasst vor allem zwei Bereiche: Wortformen, die gemäß der textkritischen Notation zu tilgen sind, sind entfernt worden. Hierarchisch untergeordnete Lemmata sind den Oberlemmata zugewiesen worden, sodass die Anzahl der Lemmata in dieser Untersuchung kleiner ist als in der zitierten. Ferner ist hier die Clusteranalyse einer PCA nachgeschaltet, sodass nur die kondensierte Information der ersten Dimensionen einbezogen wird, während die zitierte Clusteranalyse diese Datenaufbereitung nicht vorgenommen hat. Schließlich sind in der zitierten Arbeit alle Lemmata eingeflossen, während hier die seltenen ausgeschieden sind. Trotz dieser methodischen Unterschiede finden sich in den Analysen viele Gemeinsamkeiten: Sinuhe, Kemit und Nilhymnus bilden eigene Cluster. Die Textzeugen der Loyalistischen Lehre sind in der Nähe der Textzeugen der

Lehre eines Mannes für seinen Sohn. Der in der zitierten Analyse aufgezeigte Zusammenhang zwischen Bauer, Merikare und Ptahhotep wird hier ebenso bestätigt wie die Verbindung zwischen Ipuwer und Neferti. Insofern stützen sich die Clusteranalysen gegenseitig und plausibilisieren die hier vorgeschlagene Unterteilung.

3.2 Charakteristika der Cluster

Wie bei der PCA kann man in der Clusteranalyse bestimmen, welche Variablen typisch für einen Cluster sind und welche Variablen signifikant seltener verwendet werden.

Der Nilhymnuscluster ist durch die häufige Verwendung von *H^cpj*, „Hapi“ (ID:650066), *wbn*, „aufgehen; überquellen“ (ID:854500), *hrp*, „leiten; beaufsichtigen“ (ID:120150), *nhm*, „jauchzen; jubeln“ (ID:85580), *s^cnh*, „beleben; versorgen“ (ID:128910), *dw^{s.t.}*, „Duat; Unterwelt; Krypta; Grabkammer; Höhle; Abgrund“ (ID:854583), *kf^s*, „entblößen; entfernen“ (ID:164200), *š^{b.w.}*, „Geheimnis; Verborgenes; Mysterium“ (ID:158120), *mnmn.t.*, „Herde; Vieh“ (ID:70730) und *nb*, „jeder; alle; irgendein“ (ID:81660) und die geringe Verwendung von *jw*, „[aux.]“ (ID:21881), *dd*, „sagen; mitteilen“ (ID:185810), *n.j*, „gehörig zu“ (ID:850787), *r*, „[Präposition]“ (ID:91900), *n*, „[Negationspartikel]“ (ID:850806), *=j*, „[Suffix Pron. sg.1.c.]“ (ID:10030), *jm*, „da; dort; davon; dadurch; damit“ (ID:24640), *č*, „Portion; Stück“ (ID:34360), *pw*, „[Dem Pron. sg.m.]“ (ID:851517) und *jb*, „Herz; Verstand; Charakter; Wunsch“ (ID:23290) charakterisiert.

Der Kemitcluster zeichnet sich durch die hohe Verwendung von *nn-n*, „diese [Dem. Pron. pl. c.]“ (ID:500001), *zj*, „gehen“ (ID:127740), *nb-W^{s.t.}*, „Herr von Theben (Month, u. a.)“ (ID:850777), *Mnt.w*, „Month“ (ID:71660), *rmj*, „weinen; beweinen“ (ID:94180), *wr.t*, „sehr“ (ID:450161), *cnt.w*, „Myrrhenharz (für Salböl und Weihrauch)“ (ID:39010), *hnt.j*, „Vorhalle“ (ID:119070), *b³k-jm*, „der Diener da (Selbstbezeichnung des Sprechers)“ (ID:550024) und *mrj*, „lieben; wünschen“ (ID:72470) und die geringe Verwendung von *nn*, „[Negationspartikel]“ (ID:851961), *pw*, „[Dem Pron. sg.m.]“ (ID:851517), *=f*, „[Suffix Pron. sg.3.m.]“ (ID:10050) und *b³*, „Erde; Land (als Element des Kosmos); Land (geogr.-polit.); Ägypten; Erdreich (stofflich); Ackerboden; bebaubares Land; [ein Flächenmaß]“ (ID:854573) aus. Da für die Konstruktion der Cluster die ersten Dimensionen der PCA einbezogen werden, überrascht es nicht, dass hier auch die Wörter wiederzufinden sind, die auch schon in der PCA als charakteristisch für die Kemitgruppe eingestuft worden sind. Aber es werden auch Wörter als signifikant eingestuft, die in der PCA keine Rolle gespielt haben: *nb-W^{s.t.}*, „Herr von Theben (Month, u. a.)“ (ID:850777), *Mnt.w*, „Month“ (ID:71660), *rmj*, „weinen; beweinen“ (ID:94180) und *wr.t*, „sehr“ (ID:450161).

Der Cluster, der vor allem Textzeugen der Lehre eines Mannes für seinen Sohn beinhaltet, weist signifikant hohe Werte für *nb*, „Herr; Besitzer (von etwas)“ (ID:81650), *qd*, „Wesen; Gestalt; Charakter“ (ID:162430), *b³.w.*, „Ruhm; Ba-Macht“ (ID:53300), *jm.j*, „befindlich in (lokal); befindlich in (temporal); darunter (soziativ); seiend als“ (ID:25130), *pw*, „[Dem Pron. sg.m.]“ (ID:851517), *čh^c.w*, „Lebenszeit; Zeit“ (ID:40480), *ns*, „Zunge“ (ID:87800), *mrw.t*, „Liebe; Beliebtheit; Wunsch“ (ID:72650), *=tn*, „[Suffix Pron. pl.2.c.]“ (ID:10130) und *nswt*, „König von Oberägypten; König; König (Thronname der Königstitulatur)“ (ID:88040) und signifikant niedrige Werte für *hn^c*, „[Präposition]“ (ID:850800), *jr*, „[Partikel (nachgestellt zur Betonung)]“ (ID:28170), *jb.t*, „Sache; Besitz; Speisen; Opfer; Reliquie“ (ID:30750), *wnm*, „existieren; sein“ (ID:46050), *prj*, „herauskommen; herausgehen“ (ID:60920), *r-s³*, „[Präposition]“ (ID:851453), *m*, „siehe!; [Partikel]“ (ID:64440), *jnj*, „bringen; holen“ (ID:26870), *Km.t*, „Das Schwarze Land (Ägypten)“ (ID:164430) und *rm*, „Fisch (allg.)“ (ID:94160) auf. Wie auch bei der PCA fällt es auf, dass die Gruppe der signifikant seltenen Variablen zum großen Teil aus hochfrequenten Wörtern besteht.

Das Sinuhecluster verwendet signifikant oft die folgenden Wörter: *Z³-nh.t*, „Za-nehet (Sinuhe)“ (ID:705989), *m-h³.w*, „in der Nähe von (lokal); zur Zeit von (temporal)“ (ID:65060), *st.tjw*, „Asiaten“ (ID:149130), *m-hnt.yt*, „strom-auf; südwärts“ (ID:850635), *3h.t*, „Horizont; Lichtland“ (ID:227), *šms.w*, „Gefolgsmann“ (ID:856136), *Shtp-jb-R^cw*, „[Thronname Amenemhets I.]; [Thronname Sehetepibres]; [Thronname Petubastis III.]“ (ID:400449), *mš^c*, „Truppe; Heer“ (ID:76300), *b³wj*, „Abend; Nacht“ (ID:113750) und *Z-n-wsr.t*, „Sesostris“ (ID:400051). Es ist überraschend, dass es keine Überschneidung bei den Wörtern gibt, die in der PCA und in der Clusteranalyse die Sinuhegruppe auszeichnen, zumal die in der Clusteranalyse ermittelten Wörter einen hohen Einfluss auf die Konstruktion der dritten Dimension der PCA haben, also der Dimension, in der die Sinuhegruppe in der PCA als eigene Gruppierung erkennbar ist. Es gibt in diesem Cluster keine Wörter, deren signifikant niedrige Verwendung charakteristisch ist.

Der Cheticcluster weist hohe Werte für *=st*, „[Suffix Pron. sg.3.f.]“ (ID:851173), *jw.t*, „Funktion; Amt; Würde“ (ID:20430), *m.j.y*, „ebenso“ (ID:68170), *dj.w*, „Leinenstoffballen; Stoffballen“ (ID:177680), *m-hr*, „angesichts; vor“

(ID:65170), *fn*, „schwach sein“ (ID:63810), *šm.t*, „Schritt; Gang; Geschäft“ (ID:154400), *=fj*, „[Suffix Pron. sg.3.m. nach Dual]“ (ID:10060), *db.t*, „Ziegel; Platte; Barren (als Maß)“ (ID:183120) und *šbn*, „mischen; sich mischen (unter)“ (ID:153490) und niedrige Werte für *t3*, „Erde; Land (als Element des Kosmos); Land (geogr.-polit.); Ägypten; Erdreich (stofflich); Ackerboden; bebaubares Land; [ein Flächenmaß]“ (ID:854573), *=sn*, „[Suffix Pron. pl.3.c.]“ (ID:10100), *zp*, „Fall; Angelegenheit; Wesen; Mal; Heilmittel, Mittel“ (ID:854543), *pw*, „[Dem Pron. sg.m.]“ (ID:851517), *nb*, „Herr; Besitzer (von etwas)“ (ID:81650), *rnp.t*, „Jahr“ (ID:94920), *mh*, „voll sein; füllen; fassen; packen; auslegen (mit Steinen)“ (ID:854514), *hpr (m-sj)*, „jmd. verfolgen“ (ID:858535) und *st*, „es [Enkl. Pron. sg.3.c.]“ (ID:147350) auf. Wenn man nun diese Wörter mit denen aus der PCA vergleicht, fällt besonders der Zusammenhang mit den Sinuhetextzeugen auf: Während sich der Cheticcluster durch die hohe Verwendung von *dj.w*, „Leinenstoffballen; Stoffballen“ (ID:177680) und *=fj*, „[Suffix Pron. sg.3.m. nach Dual]“ (ID:10060) auszeichnen, ist es die niedrige Verwendung dieser Wörter bei der Gruppe der Sinuhetextzeugen, die die PCA als charakteristisch ermittelt hat. Die geringe Anzahl der Belege für *t3*, „Erde; Land (als Element des Kosmos); Land (geogr.-polit.); Ägypten; Erdreich (stofflich); Ackerboden; bebaubares Land; [ein Flächenmaß]“ (ID:854573) beim Cheticcluster steht der hohen Anzahl der Belege der Sinuhetextzeugen aus der PCA gegenüber. Die typischen Wörter resp. der typische Nichtgebrauch von Wörtern des Cheticclusters charakterisieren Cheti damit quasi als Antisinuhe.

Auch für den großen Cluster, der so unterschiedliche Texte wie den Bauern, den Schiffbrüchigen oder die Lehre des Ptahhotep umfasst, können die Wörter angegeben werden, die signifikant häufig bzw. selten verwendet werden. Es handelt sich um: *m*, „siehe!; [Partikel]“ (ID:64440), *nfr*, „gut sein; schön sein; vollkommen sein; vollendet sein; schön machen“ (ID:854519), *hnms*, „Freund“ (ID:118260), *sf*, „das Gestern“ (ID:133440), *hr-dl*, „sagend; mit den Worten“ (ID:650039), *hm*, „nicht wissen; negieren“ (ID:116910), *dr*, „entfernen; vertreiben; vertilgen“ (ID:180130), *jw.tj*, „welcher nicht (neg. Rel.Pron)“ (ID:22030), *jr*, „[Partikel (nachgestellt zur Betonung)]“ (ID:28170) und *q3*, „richtig sein; richtig machen“ (ID:41310), bzw. um *qm3*, „werfen; schaffen; erzeugen; ersinnen; (Metall) hämmern; treiben“ (ID:854564), *h3s.t*, „Bergland; Fremdland; Wüste“ (ID:114300), *wr*, „groß; viel; reich; bedeutend“ (ID:47271), *nb*, „jeder; alle; irgendein“ (ID:81660), *htp*, „zufrieden sein; ruhen; untergehen; zufriedenstellen; sich sättigen“ (ID:111230), *hrw*, „Stimme; Geräusch; Zank; Krach“ (ID:120010), *dj.w*, „Leinenstoffballen; Stoffballen“ (ID:177680), *jm.j*, „befindlich in (lokal); befindlich in (temporal); darunter (soziativ); seiend als“ (ID:25130), *hr*, „also; aber; denn; und danach (Konjunktion); [Partikel]“ (ID:119600) und *=fj*, „[Suffix Pron. sg.3.m. nach Dual]“ (ID:10060).

Es fällt bei den charakteristischen Wörtern auf, dass es sich bei denen, die besonders häufig in einem Cluster verwendet werden, zumeist um autosemantische Wörter handelt, während bei den Wörtern, die in einem Cluster signifikant niedrige Werte aufweisen, die synsemantischen Wörter in der Mehrzahl sind.

Wie auch bei der Gruppenbildung in der PCA sind manche Wörter mehrfach für die Herausbildung von Clustern verantwortlich, was in der folgenden Übersicht zusammengestellt wird:

Tab. 2: Charakteristische Wörter und ihr Einfluss auf die Clusterbildung.

Nilhymnus	Kemit	Lehre eines Mannes	Sinuhe	Cheti	Rest
<i>=fj</i>				+	-
<i>jm.j</i>		+			-
<i>jr</i>		-			+
<i>pw</i>	-	-	+		
<i>m</i>		-			+
<i>nb, Herr</i>		+		-	
<i>nb, jeder</i>	+				-
<i>t3</i>		-		-	
<i>dj.w</i>				+	-

Es ist nicht überraschend, in dieser Tabelle Einträge wiederzufinden, die auch schon in der PCA-Übersicht aufgetreten sind, auch wenn es sich hier nicht um eine echte Teilmenge handelt. Auch die Werte der in beiden Tabellen erscheinenden Wörter müssen nicht notwendigerweise identisch sein. So ist das Demonstrativum *pw* signifikant für die Herausbildung von Gruppen in der PCA: Der signifikant niedrigen Verwendung in der Kemit steht die signifikant

hohe in der Lehre eines Mannes für seinen Sohn gegenüber. Zusätzlich wird in der Clusteranalyse der Gebrauch in Cheti und im Nilhymnus als signifikant niedrig angesehen. Die Werte der beiden Tabellen widersprechen sich nicht, aber es kann vorkommen, dass in der Clustertabelle einzelne Cluster zusätzlich als signifikant für bestimmte Wörter angesehen werden.

Die Wörter selbst sind zumeist synsemantisch oder hochfrequent. Ein *dj.w* sticht hier heraus. Ebenfalls bemerkenswert ist, dass Sinuhe mit keinem anderen Cluster die charakteristischen Wörter teilt.

3.3 Gründe für die Clusterbildung

Die Kriterien, die dafür verantwortlich sind, dass manche Texte einen gemeinsamen Cluster bilden, sind oben herausgestellt. Es handelt sich um bestimmte verwendete Wörter. Aber den gestalterischen Willen hinter der Clusterbildung zu entdecken wird durch die Unterschiedlichkeit der Texte im größten Cluster erschwert. Direkt verständlich ist es, dass Nilhymnus und Kemit je eigene Cluster bilden, da diese Werke in ihrer Gattung von den übrigen Texten zu scheiden sind. Sinuhe beweist seine Sonderstellung innerhalb der Literatur, indem seine Textzeugen einen eigenen Cluster bilden. Wenn man schon bei den Clustern der Kemit und des Nilhymnus mit der Gattung argumentieren kann, könnte man bei Sinuhe anführen, dass sich in diesem Text verschiedene Gattungen vereinigen²⁰ und somit ein einzigartiges Konglomerat bilden. Bei den beiden großen Clustern aber versagt die Argumentationshilfe Gattung. Der Cluster um die Lehre eines Mannes für seinen Sohn beinhaltet weisheitliche Texte. Jedoch finden sich andere Lebenslehren auch in dem größten Cluster, und die zum Cluster der Lehre eines Mannes für seinen Sohn gehörende Text Sporting King ist nicht den weisheitlichen Texten zuzurechnen. So können auch weitere Faktoren für die Clusterbildung eine Rolle spielen. Es könnten regionale Parameter am Werk sein. Jedoch ist es schwer, verlässliche Daten über den Entstehungsort eines Textes, der sich nicht mit dem Fundort des Textzeugen decken muss, zu gewinnen. Evtl. könnte die Diachronie relevant sein. Der Cluster der Lehre eines Mannes für seinen Sohn umfasst Texte, die erst nach dem Mittleren Reich überliefert sind. Die Ausnahme stellt die Stele des Sehetep-ib-Re dar, die aber dadurch hier erscheint, weil die übrigen Textzeugen dieser Lehre auch in diesem Cluster vereint sind. Man könnte also argumentieren, dass die Lehre eines Mannes für seinen Sohn, die Langfassung der Loyalistischen Lehre und Sporting King erst nach dem MR entstanden sind, während die Texte des größten Clusters im MR verfasst wurden. Diese sehr hypothetische Interpretation passt sich nicht in die bisherigen Linien der Diskussion um die Datierung der literarischen Werke ein. Die Texte, die nach dem MR überliefert sind, gelten entweder allesamt als alt, aber nur spät überliefert oder allesamt als jung, d. h. nach dem MR verfasst.²¹ Eine Mischposition, dass demnach einige Texte, die nach dem MR überliefert sind, schon im MR oder früher verfasst wurden, aber einige auch nach dem MR komponiert wurden, wird nicht vertreten. Dass die Trennung des Clusters der Lehre eines Mannes für seinen Sohn von dem größten Cluster wirklich diachrone Gründe besitzt, ist weit davon entfernt, als gesichert zu gelten, und es sollte allein auf Grundlage der Clusterbildung auch keineswegs die Spätdatierung der besagten Texte propagiert werden, aber die Diskussion um die Datierung der literarischen Texte benötigt dringend neue Argumente. So schreibt Eyre m. E. zurecht in seiner Rezension des Versuchs von Stauder, alle spät überlieferten Texte als spät komponiert zu verstehen: „The important point for a reviewer to make is, therefore, that the datings suggested in the final listing (pp. 508–513) must be taken as the author’s personal proposals, tentative at best, and not securely demonstrated by firm criteria from the linguistic data-base.“²² Insofern wäre es sinnvoll, statistisch gesicherte Erkenntnisse über Text-nähe und -ferne in die Datierungsfrage einfließen zu lassen.

²⁰ Morenz 1997, 1: „Als Grundform wurde die fiktive Autobiographie, in die erstaunlich viele andere Textgattungen (z. B. Hymnus, Gebet, Brief usw.) eingebettet wurden [...], gewählt.“

²¹ Vgl. Moers et al. 2013.

²² Eyre 2014, 326.

4 Fazit

Abschließend seien die Unterschiede zwischen der hier vorgestellten Clusteranalyse und der bisher vorgeschlagenen Klassifizierungen der ägyptischen Literatur genannt: Oben wurde auf die Diskrepanz zwischen dem abstrakten Klassifikationsmodell und der Einsortierung konkreter Werke hingewiesen. Texte lassen sich schwer einer oder genau einer Gattung zuweisen, sodass die Verortung eines Textes in einem gattungsorientierten Klassifikationsmodell erschwert ist. Ferner sind die Kriterien der Zuweisung nicht in dem Maße offengelegt, dass eine Falsifizierung ermöglicht wird. All diese Probleme umgeht die Clusteranalyse. Sie bringt die Kriterien, d. h. in diesem Fall die Wörter zu Tage, die bestimmte Texte zu einem Cluster verbinden. Gerade weil auf ein präkombiniertes Klassifikationsschema verzichtet wird, bereiten die Texte, die mehrere unterschiedliche Gattungsmerkmale aufweisen, bei einer Gruppenbildung keinerlei Probleme. Schließlich werden alle nicht seltenen Wörter gleichermaßen in die Analyse einbezogen, sodass latente Strukturen aufgedeckt werden können. Der Gefahr, mit nicht balancierten Kriterien unzutreffende Ergebnisse zu erzielen, ist damit die Grundlage entzogen. Für ein allgemeines Klassifikationsmodell ägyptischer Literatur scheint somit eine Clusteranalyse den Modellen überlegen, die Texte in eine Gattungshierarchie eingliedern wollen.

Bibliographie

- Assmann (1996): Jan Assmann, „Kulturelle und literarische Texte“. In: Antonio Loprieno (Hrsg.), *Ancient Egyptian Literature – History and Forms* (Probleme der Ägyptologie 10), Leiden, Köln und New York, 59–81.
- Berlin-Brandenburgische Akademie der Wissenschaften (2004-): Berlin-Brandenburgische Akademie der Wissenschaften, *Thesaurus Linguae Aegyptiae*. <http://aaew.bbaw.de/tla/> (abgerufen am 24. 06. 2020).
- Berlin-Brandenburgische Akademie der Wissenschaften (2018): Berlin-Brandenburgische Akademie der Wissenschaften, *Teilauszug der Datenbank des Vorhabens „Strukturen und Transformationen des Wortschatzes der ägyptischen Sprache“ vom Januar 2018*. urn:nbn:de:kobv:b4-opus4-29190 (abgerufen am 24. 06. 2020).
- Blumenthal (1998): Elke Blumenthal, „Prolegomena zu einer Klassifizierung der ägyptischen Literatur“. In: Christopher J. Eyre (Hrsg.), *Proceedings of the Seventh International Congress of Egyptologists* (Orientalia Lovaniensia Analecta 82), Leuven, 173–183.
- Burkard/Thissen (2003): Günter Burkard und Heinz-Josef Thissen, *Einführung in die altägyptische Literaturgeschichte I. Altes und Mittleres Reich* (Einführungen und Quellentexte zur Ägyptologie 1), Münster, Hamburg und London.
- Burrows (2002): John Burrows, „Delta‘. A Measure of Stylistic Difference and a Guide to Likely Authorship“. In: *Literary and Linguistic Computing* 17, 267–287.
- Büttner et al. (2017): Andreas Büttner, Friedrich Michael Dimpel, Stefan Evert, Fotis Jannidis, Steffen Pielström, Thomas Proisl und Isabella Reger, „»Delta« in der stilometrischen Autorschaftsattribution“. In: *Zeitschrift für digitale Geisteswissenschaften*. text/html Format. DOI: 10.17175/2017_006 (abgerufen am 24. 06. 2020).
- Eyre (2014): Christopher Eyre, „Rezension zu Andréas Stauder, *Linguistic Dating of Middle Egyptian Literary Texts*“. In: *Lingua Aegyptiae* 22, 321–326.
- Husson et al. (2017): François Husson, Sébastien Lê und Jérôme Pagès, *Exploratory Multivariate Analysis by Example Using R*, Milton.
- Lê et al. (2008): Sébastien Lê, Julie Josse und François Husson, „FactoMineR. An R Package for Multivariate Analysis“. In: *Journal of Statistical Software* 25, 1–18.
- Moers et al. (2013): Gerald Moers, Kai Widmaier, Antonia Giewekemeyer, Arndt Lümers und Ralf Ernst (Hrsg.), „*Dating Egyptian Literary Texts*“ Göttingen, 9–12 June 2010 (Lingua Aegyptia Studia Monographica 11), Hamburg.
- Morenz (1997): Ludwig D. Morenz, „Kanaanäisches Lokalkolorit in der Sinuhe-Erzählung, insbesondere: Sinuhe als (Unter)Patriarch von Ammu-nanśis Gnaden in der (Beinahe-) Dublette Sinuhe B 78 und B 107f.“. In: *Zeitschrift des Deutschen Palästinavereins* 113, 1–18.
- Parkinson (1996): Richard Parkinson, „Types of Literature in the Middle Kingdom“. In: Antonio Loprieno (Hrsg.), *Ancient Egyptian Literature – History and Forms* (Probleme der Ägyptologie 10), Leiden, Köln und New York, 297–312.
- Reinsel et al. (2018): David Reinsel, John Gantz und John Rydning, *Data Age 2025. The Digitization of the World from Edge to Core*. <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> (abgerufen am 24. 06. 2020).
- Schweitzer (2013): Simon D. Schweitzer, „Dating Egyptian Literary Texts: Lexical Approaches“. In: Gerald Moers, Kai Widmaier, Antonia Giewekemeyer, Arndt Lümers und Ralf Ernst (Hrsg.), „*Dating Egyptian Literary Texts*“ Göttingen, 9–12 June 2010 (Lingua Aegyptia Studia Monographica 11), Hamburg, 177–190.
- Schweitzer (2019a): Simon D. Schweitzer, *simondschweitzer/aed-tei: AED-TEI Version 1.0 (Version v1.0) [Data set]*. Zenodo. DOI:10.5281/zenodo.3580939 (abgerufen am 24. 06. 2020).
- Schweitzer (2019b): Simon D. Schweitzer, *Ancient Egyptian Dictionary*. <https://simondschweitzer.github.io/aed/index.html> (abgerufen am 24. 06. 2020).

- Schweitzer (2019c): Simon D. Schweitzer, *simondschweitzer/aed: AED – Ancient Egyptian Dictionary Version 1.0 (Version v1.0) [Data set]*. Zenodo. DOI:10.5281/zenodo.3581069 (abgerufen am 24. 06. 2020).
- Schweitzer (2020): Simon D. Schweitzer, *Word Frequency: Middle Egyptian Literary Texts (Version 1.0) [Data set]*. Zenodo. DOI 10.5281/zenodo.3924222 (abgerufen am 30. 06. 2020).
- Seidlmayer (1990): Stephan J. Seidlmayer, *Gräberfelder aus dem Übergang vom Alten zum Mittleren Reich. Studien zur Archäologie der Ersten Zwischenzeit* (Studien zur Archäologie und Geschichte Altägyptens 1), Heidelberg.
- Seidlmayer/Hafemann (2011): Stephan J. Seidlmayer und Ingelore Hafemann, *Handbuch zur Benutzung des Thesaurus Linguae Aegyptiae (TLA). Auf der Grundlage der Hilfetexte des Thesaurus Linguae Aegyptiae (TLA), erarbeitet von Stephan J. Seidlmayer, überarbeitet und zusammengestellt von Ingelore Hafemann*, Berlin http://aaew.bbaw.de/hgl/0/Manual_2012_02_02.pdf (abgerufen am 24. 06. 2020).

Software

- FactoMineR <http://factominer.free.fr/> (abgerufen am 24.06.2020).
- Factoshiny <http://factominer.free.fr/graphs/factoshiny.html> (abgerufen am 24.06.2020).
- R <https://www.r-project.org/> (abgerufen am 24.06.2020).

Abbildungsverzeichnis

- Abbildung 1: PCA mit allen Wörtern, erste und zweite Dimension.
©Simon D. Schweitzer, CC-0
- Abbildung 2: Modifizierte PCA ohne die seltenen Wörter, erste und zweite Dimension.
©Simon D. Schweitzer, CC-0
- Abbildung 3: Modifizierte PCA ohne die seltenen Wörter, erste und zweite Dimension, Textzeugen der Lehre des Ptahhotep.
©Simon D. Schweitzer, CC-0
- Abbildung 4: Modifizierte PCA ohne die seltenen Wörter, erste und zweite Dimension, Textzeugen der Lehre für Merikare.
©Simon D. Schweitzer, CC-0
- Abbildung 5: Modifizierte PCA ohne die seltenen Wörter, dritte und vierte Dimension.
©Simon D. Schweitzer, CC-0
- Abbildung 6: Dendrogramm der Clusteranalyse.
©Simon D. Schweitzer, CC-0

