

Theoria cum praxi

Reinhold Scholl

Das Motto des Universalgelehrten Gottfried Wilhelm Leibniz, der als einer der Väter der Digital Humanities gelten kann, weil er die Voraussetzungen für die binäre Welt der Einser und Nullen mitgeschaffen hat, gilt auch für diese Zeitschrift.

Die in diesem Heft versammelten Beiträge eint, daß sie es mit der Anwendung der Arbeitsweisen der Digital Humanities auf bestimmte Fragestellungen zu tun haben. Das geht von der Forderung nach einer sinnvollen und nachvollziehbaren Erstellung von Stoppwörtern, über die händische Codierung von Texten, um sie über den Umweg der Maschinenlesbarkeit wieder den Menschen verständlich zu machen, der Auswertung einer personenbezogenen Datenbank im unteren Donauraum in griechisch-römischer Zeit auf der Basis von Inschriften bis hin zu einer textminigbasierten Wortanalyse zur Klärung einer Ernährungsfrage in Athen in klassischer Zeit.

Der Beitrag von Patrick J. Burns beschäftigt sich mit dem Erstellen von sogenannten Stoppwörtern. Alle Tools, die die automatische Bearbeitung von größeren Textcorpora zum Ziel bzw. als Hilfsmittel haben, waren von Anbeginn an daran interessiert, um Rechen- und Rechnerkapazität zu sparen sowie um aussagefähige Suchergebnisse zu erhalten, Stoppwörter aus ihren Analysen heraus zu nehmen. Burns bemängelt in seinem Artikel u.a., daß die Kriterien für diese Listen oft nicht transparent gemacht, nicht immer eingehalten und auch nicht dokumentiert wurden. Oft kommt zu den Algorithmen noch eine subjektive Bearbeitung hinzu. Zu beachten ist ebenfalls, daß die Stoppwörterlisten, auch wenn Sie nach klaren Algorithmen gebildet werden, immer abhängig sind von dem jeweils zugrundeliegenden Textcorpus, aus dem sie gewonnen werden, nämlich von dessen zeitlicher Erstreckung und von ihrem Inhalt. Diesen undurchschaubaren Stoppwörterlisten für einzelne Sprachen und besonders für Latein stellt er seinen Beitrag gegenüber. Sein "paper describes the development of a generalizable method for building stoplists in the Classical Language Toolkit (CLTK), an open-source Python platform for natural language processing research on historical languages." Das Plädoyer für mehr Transparenz und auch Dokumentation kann man nur unterstützen. Andererseits stellt sich einem mehr geisteswissenschaftlich vorgebildeten interessierten IT-Laien die Frage, ob bei der heutigen Kapazität der Rechner Beschränkungen wie die Herausnahme von Stoppwörtern in negativen oder positiven Stoppwörterlisten noch notwendig sind oder besser, ob man es nicht lieber den Anwenderinnen und Anwendern überlassen sollte, ob sie dieses Tool der Stoppwörterlisten nutzen wollen, zumal es von der Fragestellung abhängig ist, welche Wörter zur Bedeutung und Deutung beitragen. Stoppwörter erfüllen zwar per definitionem in erster Linie grammatikalische und syntaktische Funktionen und tragen weniger zum Inhalt, wohl aber zum Sinn und zur Bedeutung bei. Das versteht sich z.B. bei Verneinungen von selbst und Konjunktionen können komplexe Zusammenhänge verdeutlichen. Bei einer Untersuchung zum Sprachstil (z.B. Parataxe-Hypotaxe, Litotes, usw.) sind diese Stoppwörter zumindest für die Statistik nicht irrelevant. Strenggenommen müßte für jede Untersuchung eine eigene Stoppwörterliste erstellt werden.

Bei dem Beitrag „Digital Fabius Pictor“ von Kevin Strassburger könnte man ketzerisch die Frage stellen, ob man wirklich einen Mehrwert für die Leserinnen und Leser generiert, wenn man den Text einer analogen Ausgabe, in der sich die Schriftsetzer alle Mühe gegeben haben, die unterschiedlichen Informationen auch drucktechnisch anschaulich zur Verfügung zu stellen, wie z.B. die verschiedenen Schriften (griechische und lateinische Buchstaben), Markierungen mit Linien, um den Originaltext von der Übersetzung zu trennen, wie in der Ausgabe von Müller geschehen, in eine digitale Form bringt oder Peters Ausgabe, der statt der Übersetzung einen zweigeteilten Apparat bietet, der unter dem griechischen Originaltext durch Absatzstriche voneinander getrennt einen lateinischsprachigen Kommentar und davon wiederum getrennt einen textkritischen Apparat mit all den für Nichtphilologen geheimnisvollen bzw. kryptisch anmutenden Abkürzungen, die bisweilen Anfänger zum Verzweifeln bringen können, bietet, digitalisiert. Die Abkürzungen in diesen Ausgaben, die in gewisser Weise Herrschaftswissen symbolisieren, sind u.a. verwendet worden, um Platz- und Druckkosten zu sparen. Diese hätte man im Zuge der Digitalisierung auflösen und mit einem gesonderten `Tag` versehen können, um das Original von der Bearbeitung zu unterscheiden. Ein weiterer Schritt wäre dann die Umwandlung des maschinenlesbaren Textes wieder in einen für Menschen lesefreundlichen Text mit all den zusätzlichen Annotationen, die, wie der Verfasser auch schreibt, durchaus in unterschiedlicher Form – nach Möglichkeit sogar wahlweise – angezeigt werden könnten. Wenn dann noch alle Wörter eine morphologische Analyse erhalten und eventuell ihre syntaktische Funktion erklärt wird und nicht nur wie hier die Personennamen (antike und moderne, wobei aber kein Unterschied zwischen ihnen gemacht wird), Ortsnamen und Literatur und die mit dem Layout verbundenen Dinge annotiert sind, ist der Mehrwert einer digitalen Ausgabe sicherlich größer und es werden zukünftig bessere Suchen und auch Auswertungen ermöglicht. Dazu war diese praktische Handarbeit der Digitalisierung und deren Beschreibung ein erster Schritt.

Die Auswertung der Personendatenbank (<http://romans1by1.com>) durch Rada Varga und Imola Boda ist die Fortsetzung eines früheren Beitrages in dieser Zeitschrift (<https://journals.ub.uni-heidelberg.de/index.php/dco/article/view/35822>). Auf der Homepage der Datenbank (zuletzt eingesehen am 5.11.2018) steht: “Romans 1by1 is a population database recording people identified in Greek and Roman epigraphy. The areas integrally covered at this point are the provinces of Moesia Inferior, Moesia Superior and Dacia. From Pannonia Superior, we have integrated so far the urban centers of Poetovio, Neviodunum, Siscia, Savaria, Brigetio and Scarbantia. Romans 1by1 is a dynamic tool and can be considered a constant work in progress.” Diese sehr gut strukturierte und dokumentierte Datenbank liefert nun erste Auswertungsergebnisse hinsichtlich der Art der Inschriften, ihrer Sprache, der Zusammensetzung der Bevölkerung in Bezug auf Alter, Geschlecht, Rechtsstatus, Tätigkeiten in Verwaltung, Militär, Kult und Vereinswesen. Im Zuge der Personendatenbank wurde sicherlich auch mehr oder weniger ein Corpus der untersuchten Inschriften erstellt, die – sofern bereits digitalisiert – über Verlinkungen für die Nutzerinnen und Nutzer als Volltext zur Kontrolle einzusehen sind. Was spricht eigentlich dagegen, diese Inschriften zukünftig auch als Gesamtkorpus zur Verfügung zu stellen?

Der Titel des Beitrages von Sven Brandt „κρίθη oder πύρος? - Eine systematische Annäherung im klassischen Athen mit Hilfe der Kookkurrenzanalyse von eAQUA und den Word Frequency Statistics der Perseus Digital Library“ deutet bereits an, daß die beiden im Titel genannten Verfahren der Digital Humanities nicht ohne weitere „menschliche“ Zuarbeit zu hilfreichen Argumenten oder Ergebnissen bei der Beantwortung der Frage führen. Es wird deutlich gemacht, daß diese Verfahren noch nicht ganz ausgereift sind, sondern der Überprüfung und Kontrolle durch die Anwenderinnen und Anwender bedürfen. „Blended Readings“ heißt das neue Schlagwort. *Distant reading* und *close reading* (vgl. [Schubert DCO 1,1 2015](#)) werden dabei

miteinander durch klassische Diskursanalyse und Hermeneutik ergänzt. Provozierend gesagt, was man in den Altertumswissenschaften methodisch, aber noch nicht computerunterstützt, bisher fast immer getan hat: Zu einer Fragestellung, einem Problem, das man lösen wollte, hat man die entsprechenden Quellen gesammelt, gesichtet, geordnet, analysiert und ausgewertet und schließlich in adäquater Form und nachvollziehbar präsentiert. Genau das geschieht auch im Beitrag von Sven Brandt. Die Quellensammlung erfolgt über die entsprechenden Termini in den zur Verfügung stehenden Datenbanken. Die Auswahl der Autoren beschränkt sich auf ihre Wirkungszeit im 5. und 4. Jh. v. Chr. in Athen. Dazu ist das Frequenzanalysetool von Perseus ein gutes Hilfsmittel. Das Problem ist allerdings, daß dieses Tool bei Homonymen (Homographen) nicht unterscheiden kann, zu welchem Wort jeweils die gleichaussehenden Grapheme gehören. Hier muß dann der menschliche Verstand die Zuordnung vornehmen. Das gleiche Problem der Homographie/Homonymie taucht auch in der Kookkurenanalyse von eAQUA auf. Allein schon aus diesem Grund können diese beiden Tools nicht allein ohne Überprüfung und Kontrolle angewandt werden, ohne zu verfälschenden Ergebnissen zu gelangen. Textminingtools ersetzen nicht Sprachkompetenz, im Gegenteil, sie setzen diese in einem hohen Maße voraus. Das Ergebnis des Beitrages zeigt, daß in den literarischen Quellen nicht der Gegensatz Gerste oder Weizen, sondern das Gemeinsame, nämlich Weizen und Gerste betont wird. In der Tat bringt die textmininggestützte Analyse der literarischen Quellen der klassischen Zeit in Athen nur eine Annäherung – aber immerhin die – an den Komplex zur Getreidefrage in Athen (Bedarf, Anbau, Import, Konsum), aber natürlich keine Lösung. Eine weitere Annäherung könnte darin bestehen, dieses Verfahren auch auf die Inschriften anzuwenden und auch auf Autoren, die zwar später lebten, aber die klassische Zeit behandeln. Und schließlich könnte man als weitere Ergänzung die entsprechenden „Halb-“ und „Fertigprodukte“ wie ἄλφιτον (Gerstenmehl) und μᾶζα (Gerstenkuchen) auf der einen und ἄλευρον (Weizenmehl) und ἄρτος (das Brot (aus Weizen)) auf der anderen Seite mit in die Untersuchung aufnehmen. Denn die meisten Stadtbewohner Athens kamen lediglich mit diesen Halb- und Fertigprodukten in Kontakt und für deren Herstellung gab es die στροποιοί (Müller/Bäcker). Auch diese Vorgehensweise ist ein Beispiel für die Sinnhaftigkeit einer ganzheitlichen Betrachtungsweise.

Alle Beiträge dieses Heftes zeigen anschaulich und auf unterschiedliche Art und Weise, daß die Zuhilfenahme von Verfahren und Arbeitsweisen der Digital Humanities den Classics helfen können.

Autorenkontakt¹

Prof. Dr. Reinhold Scholl

Universität Leipzig
Historisches Seminar, Alte Geschichte
Beethovenstr. 15
04107 Leipzig

Email: scholl@uni-leipzig.de

¹ Die Rechte für Inhalt, Texte, Graphiken und Abbildungen liegen, wenn nicht anders vermerkt, bei den Autoren. Alle Inhalte dieses Beitrages unterstehen, soweit nicht anders gekennzeichnet, der Lizenz CC BY 4.0.