

**Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig**

**Text Mining am Beispiel der Dramen
Shakespeares. Welche neuen Erkenntnisse
können moderne formale Methoden liefern?**

by

Thomas Efer, Gerhard Heyer, and Jürgen Jost

Preprint no.: 99

2014



THOMAS EFER, GERHARD HEYER, JÜRGEN JOST

**Text Mining am Beispiel der Dramen Shakespeares.
Welche neuen Erkenntnisse können moderne formale
Methoden liefern?**

I. Einleitung

Warum sind Shakespeares Werke so einzigartig? Was zeichnet sie aus gegenüber den Werken seiner Zeitgenossen? Wie entwickeln sich Vokabular, Schreibstil und Dramenstruktur über seine Schaffenszeit hinweg? Dies sind natürlich grundlegende Fragen für die literarische und wissenschaftliche Beschäftigung mit Shakespeare. Aber was können Informatiker und Mathematiker dazu beitragen?

In diesem Beitrag wollen wir Methoden der Informatik und Konzepte der Mathematik zur formalen Untersuchung der Sprach- und Handlungsstruktur in den Dramen Shakespeares entwickeln, vorstellen und anwenden. Wir wollen dabei sowohl theoretische Konzepte herausarbeiten als auch neue praktische Verfahren einsetzen. Wir hoffen, damit auch der Literaturwissenschaft neue Anregungen geben und Fragestellungen eröffnen zu können, und dass dies letztlich zu vertieften Einsichten in den sprachlichen und dramatischen Aufbau der Werke Shakespeares führt und die literarische Entwicklung im Schaffen Shakespeares auch durch formale Kriterien belegen und Unterschiede zwischen Shakespeare und anderen Dramatikern anhand quantitativer Kriterien aufweisen kann.

Methodisch bewegen wir uns dabei also im Gebiet der eHumanities, also der geisteswissenschaftlich orientierten Teildisziplin der Informatik, und genauer im Bereich des Text Mining.¹ Text Mining untersucht Datenstrukturen unter Berücksichtigung der Besonderheiten textueller Daten (zum Beispiel Wortverteilungen nach dem Zipf'schen Gesetz) mit formalen Verfahren und nach formalen Kriterien. Durch einen solchen Einsatz formaler Verfahren und Kriterien lassen sich sowohl quantitative Maße zum Vergleich verschiedener Texte oder handlungstragender

1 Zu vielen der von uns benutzten informatischen Fachbegriffe gibt es noch keine etablierte und akzeptierte deutsche Übersetzung. Die sprachschöpferische Kraft des Englischen ist seit den Tagen Shakespeares ungebrochen, während die sprachlichen Ausdrucksmöglichkeiten von Schlegel, Tieck und den anderen seinerzeitigen Shakespeareübersetzern kein modernes Äquivalent im Deutschen zu besitzen scheinen.

oder sprachlicher Strukturen innerhalb eines Textes gewinnen als auch Eigentümlichkeiten von Texten oder bestimmter struktureller Merkmale entdecken.

Bemerkenswerterweise wurden digitale Werkzeuge für die Geisteswissenschaften schon in den Anfangsjahren des Computers entwickelt. Roberto Busa begann schon in der Mitte der 1940er Jahre, in Kooperation mit IBM an einem automatisch erstellten lemmatisierten Suchindex zu den Werken des Thomas von Aquin zu arbeiten. In Deutschland entwickelte der Physiker Wilhelm Fucks seit den 50er Jahren statistische Verfahren zum systematischen formalen Vergleich von Texten und damit auch von deren Autoren.² Die heutzutage fast unbegrenzt verfügbare Rechenleistung eröffnet den digitalen Geisteswissenschaften viele neue Möglichkeiten, bedarf aber natürlich der Fokussierung durch tragfähige inhaltliche Konzepte und der Lenkung durch profunde wissenschaftliche Methoden. Einige dieser Methoden sind in den letzten Jahren in der Mathematik entwickelt worden, insbesondere auch neue Maße zur Bewertung der Komplexität von Prozessen und Strukturen. Dabei gilt etwas als komplex, wenn es weder regellos, wirr, willkürlich, zufällig ist, noch einfachen und starren Regeln folgt, sondern wenn es eine reichhaltige innere Struktur besitzt, also etwas, was die Dramen Shakespeares in ganz besonderer Weise auszeichnet. Insbesondere greifen bei Shakespeare ein außergewöhnlich umfangreiches Vokabular, eine große sprachliche Ausdruckskraft und ein subtiler Handlungsaufbau ineinander, und dies sollte dann auch durch formale und quantitative Kriterien belegt und aufgeschlüsselt werden können.

Bevor wir allerdings dazu kommen, müssen erst einmal einige Vorarbeiten geleistet werden. Dies beginnt mit der Frage nach der digitalen Verfügbarkeit der Quellen und deren Prüfung auf Brauchbarkeit zum Einsatz formaler Methoden. Als Datengrundlage haben wir schließlich zum einen die populäre Moby-Edition ausgewählt, daneben auch die Folger Digital Texts, sowie *Hamlet*-Versionen des Shakespeare Quartos Archive. Die Berücksichtigung von Quartos erlaubt das Aufzeigen der großen Varianz innerhalb der noch zu Shakespeares Lebzeiten erschienenen Ausgaben und gleichzeitig eine Überprüfung der Robustheit der verwendeten beziehungsweise neu entwickelten Verfahren gegenüber dieser Art des textuellen ›Rauschens‹. Daneben wird versucht, über jüngere Bearbeitungen einen Brückenschlag von modernsprachlich etablierten Text-Mining-Verfahren, wie etwa Stimmungsanalysen, zum ›schwierigen‹ Shakespeare'schen Englisch durchzuführen. Für die deutschen Übersetzungen und die deutschen Dramen, die wir mit den Shakespeareübersetzungen vergleichen, haben wir die Texte des Projekt Gutenberg herangezogen. Insgesamt können aber die Verfügbarkeit und Qualität der für unser

2 Zusammenfassend siehe Wilhelm Fucks: *Nach allen Regeln der Kunst*. Stuttgart 1968. Solche Verfahren werden auch routinemäßig in der Shakespeareforschung eingesetzt, siehe zum Beispiel Hugh Craig: »Authorship«. In: Arthur E. Kinney (Hrsg.): *The Oxford Handbook of Shakespeare*. Oxford 2012, S. 15–30.

Projekt relevanten maschinenlesbaren Quellen noch nicht als befriedigend angesehen werden.

Konkret wird dann in unserem Beitrag unter der Prämisse einer explorativen Herangehensweise an die Textdaten beispielhaft eine Prozesskette für ›Drama Mining‹ vorgestellt. Durch eine klare Struktur der Einzelabläufe und ihrer Verknüpfung wird dabei die zur Einschätzung der Resultate notwendige methodische Transparenz gewährleistet:

- Konvertierung der verfügbaren Dramen in ein einheitliches Format auf Basis des TEI-Schemas unter Berücksichtigung der Editionsgeschichte der einzelnen Texte
- Definition verschiedener gegebenenfalls parametrisierbarer Graph-Induktions-Verfahren zur Extraktion struktureller Merkmale (z. B. Redeabfolge, direkte Nennung von Charakteren, szenische Gliederung oder Regieanweisungen) und sekundärer Strukturen (zum Beispiel Kookkurrenzen von Personen) aus den TEI-Dateien
- Definition und Auswertung von textstatistischen, informationstheoretischen und konventionellen Maßen zur Charakterisierung der Texte mittels bestimmter Struktureinheiten (zum Beispiel Vokabulargröße einzelner Charaktere, Wortentropie in langen Monologen, verwendete Fragezeichen pro Szene)
- Formale Analyse des Aufbaus und der Handlungsstruktur anhand der Abfolge der auftretenden Personen und der Verweise in den Dia- und Monologen
- Systematische Bewertung von Komplexitäten mittels mathematisch definierter Maße
- Kombinierte Betrachtung von Strukturen und Maßen über die Dramen hinweg, und Ermittlung charakteristischer Kennwerte, mit denen sich a) ein Trend über Shakespeares gesamte Schaffenszeit oder b) eine eindeutige Abgrenzung gegenüber anderen Autoren erkennen lässt
- Abgleich der gefundenen (oder besser vermuteten) Erkenntnisse mit bisheriger Forschung

Als Resultat der Arbeiten werden neue Einsichten in den spezifischen Schreibstil und die besondere Strukturierung von Shakespeares dramatischen Werken erwartet. Einsichten, die selbstverständlich einer genauen fachwissenschaftlichen Deutung und Einordnung bedürfen. Es bleibt auch zu bemerken, dass eine maschinelle Verarbeitung literarischer Texte meist auch irgendwelche Artefakte produziert, die dann vom menschlichen Nutzer eliminiert werden müssen. Es geht also nicht darum, die Textanalyse durch Literaturwissenschaftler durch maschinelle Verfahren zu ersetzen, sondern es sollen die traditionellen Verfahren der Literaturwissenschaft durch neue Werkzeuge erweitert werden. Im Wechselspiel zwischen menschlichem Nutzer und rechnergestützten Verfahren sollen sich die jeweiligen Stärken ergänzen.

II. Literaturwissenschaftliche Einordnung

Literaturwissenschaft beschäftigt sich mit dem Verhältnis von Form und Inhalt, welches durchaus ein Spannungsverhältnis sein kann, literarischer Werke, mit der Einordnung dieser Werke in den biographischen, sozialen und kulturellen Kontext ihrer Autoren, mit der Wirkungsgeschichte und mit dem systematischen Vergleich ähnlicher oder unterschiedlicher Werke. Diese Spannweite bedingt und erfordert einen methodischen Pluralismus. Es geht ihr implizit oder explizit auch um ästhetische Fragen, die Bewertung der Qualität literarischer Werke.

Wir wollen nun die Position unseres Ansatzes im Gesamtzusammenhang der Literaturwissenschaft skizzieren. Es ist eine alte Frage, ob die formale Gestalt eines Werkes als eine absichtliche, bewusst reflektierte Gestaltung durch dessen Schöpfer anzusehen ist oder ob künstlerische Intuition auf einer eher unschweligen Ebene wirkt. Letzteres schlägt sich in Begriffen wie demjenigen des Genies nieder, und kann vielleicht auch bis zu einem gewissen Grade mit hermeneutischen Ansätzen erahnt werden, ist aber mit den traditionellen Methoden der Literaturwissenschaft schwer zu präzisieren. Mit unserem formalen Ansatz wollen wir nun gerade solche Schichten der Komplexität von Werken freilegen und einer präzisen Analyse zugänglich machen, die wohl kaum auf bewussten Konstruktionen beruhen und die sich daher bisher auch der formalen Analyse entzogen haben.

Außerdem wollen wir Verfahren zum systematischen formalen Vergleich verschiedener Texte entwickeln. Durch die von uns vorgeschlagenen und eingesetzten Methoden des ›distant reading‹ können nicht nur formale Profile verschiedener Autoren, Stile oder Perioden entwickelt werden, sondern es können auch neue Erkenntnisse über literarische Entwicklungen und Einflüsse gewonnen werden.³ Shakespeare ist dabei für unsere Zwecke besonders gut geeignet. Wir können nämlich nicht nur versuchen, die einzigartige Komplexität seiner Sprachverwendung und seines Handlungsaufbaus anhand neuer formaler Kriterien zu belegen, sondern wir können auch diese formalen Maße für seine Originaltexte mit denjenigen seiner deutschen Übersetzungen vergleichen. Wir können dann untersuchen, inwieweit sich die sprachliche Komplexität durch die Übersetzungsgeschichte hin entwickelt, von den frühen Übersetzungen von Wieland und Eschenburg über das Übersetzungswerk von August Wilhelm Schlegel und seine Weiterführung durch Dorothea Tieck und Wolf von Baudissin und die konkurrierenden Übersetzungen durch Johann Heinrich Voss und seine Söhne, die vielfältigen weiteren Übersetzungen des 19. Jahrhunderts bis zu denjenigen des 20. Jahrhunderts, wie der sprachlich eigenwilligen von Friedrich Gundolf oder der als sprachlich modern konzipierten von Erich Fried.⁴

3 Vgl. Franco Moretti: *Distant Reading*. London 2013.

4 Grundlegend zur Übersetzungsgeschichte Roger Paulin: *The Critical Reception of Shakespeare in Germany 1682–1914. Native Literature and Foreign Genius*. Hildesheim 2003

Auch hinsichtlich der Analyse der Handlungsstruktur bieten die vielen Dramen Shakespeares reiches Material. Dabei lassen sich die Interaktionsstrukturen gut durch Graphen veranschaulichen. Auf diese Weise können wir auch moderne Methoden der Visualisierung zur Darstellung von Handlungsgefügen einsetzen. Dadurch können wir nicht nur die verschiedenen Dramen miteinander vergleichen und versuchen, einerseits durchgängige Grundmuster zu identifizieren und andererseits spezifische Eigenarten des Aufbaus einzelner Dramen herauszuarbeiten, sondern wir können auch erfassen, inwieweit die einzelnen Personen in den jeweiligen Dramen handlungstragend sind und welches die charakteristischen Abfolgenmuster der Handlungsstrukturen sind. Dass hierdurch wesentliche Einsichten in die dramatische Technik Shakespeares gewonnen werden können, hat schon August Wilhelm Schlegel gesehen; er schrieb: »In einem schon vor einer Anzahl von Jahren abgefaßten Versuch über Romeo und Julia habe ich die sämtlichen Auftritte nach der Reihe durchgegangen und die innere Notwendigkeit eines jeden in Bezug auf das ganze geprüft.«⁵ Wir wollen hier solche Versuche mit neuen formalen Methoden beginnen.

Die Beziehung zwischen den von uns entwickelten und eingesetzten formalen Maßen und ästhetischen Kriterien soll in diesem Beitrag allerdings nicht thematisiert und erörtert werden.⁶

III. Zum mathematischen Formalismus

Schon Claude Shannon, der Begründer der Informationstheorie, hat 1948 in seinem grundlegenden Werk *A Mathematical Theory of Communication* die Entropierate der englischen Sprache abgeschätzt.⁷ Zunächst einmal kann man die Häufigkeitsverteilung von Buchstaben in einer Sprache oder einem Text untersuchen. Bekanntlich sind nicht alle Buchstaben gleich häufig vertreten, und im Deutschen

5 Vgl. August Wilhelm Schlegel: *Vorlesungen über dramatische Kunst und Literatur. Kritische Ausgabe*. Hrsg. von Giovanni Vittorio Amoretti. Bonn u. a. 1923. Bd. 2, S. 132.

6 Der Vorschlag des Mathematikers George D. Birkhoff: »A Mathematical Approach to Aesthetics«, *Scientia* 50 (1931), S. 133–146, war vielleicht noch etwas amateurhaft, aber Max Bense: *Einführung in die informationstheoretische Ästhetik. Grundlegung und Anwendung in der Texttheorie*. Reinbek 1969, und ders.: *Programmierung des Schönen*. Baden-Baden 1960, entwickelte ein seinerzeit einflussreiches ästhetisches Programm auf der Basis quantitativer informationstheoretischer Maße. Für weitere Referenzen vgl. Wolfgang Ebeling, Jan Freund und Frank Schweitzer: *Komplexe Strukturen. Entropie und Information*. Stuttgart 1998, dort Kapitel 6.

7 Claude Shannon: »A Mathematical Theory of Communication«, *Bell System Technical Journal* 27 (1948), S. 379–423 und S. 623–656; siehe auch Wolfgang Ebeling u. a.: *Komplexe Strukturen* (Anm. 6) für Weiterführungen dieses Ansatzes.

gilt die Häufigkeitsreihenfolge E, R, N, I, S, T, U ... Derartige Verteilungen folgen typischerweise einem Potenzgesetz. Oft spricht man von dem Zipf'schen Gesetz. Man kann dann auch analog die Häufigkeitsverteilungen von Buchstabenpaaren oder längeren Buchstabenketten feststellen. Wesentlich ist, dass die relative Häufigkeit eines Buchstabenpaares sich typischerweise von dem Produkt der Einzelhäufigkeiten der beteiligten Buchstaben unterscheidet. Shannon stellte daher die Frage, wieviel Information man aus der Kenntnis eines Blocks von n Buchstaben über die nachfolgenden Buchstaben gewinnt. Je mehr Buchstaben man kennt, umso leichter fällt es, die nachfolgenden zu erraten. Oder anders ausgedrückt, die Entropie von Buchstabenketten wächst langsamer als ihre Länge, und dies reflektiert die Struktur der Sprache.

Nun zielt dies allgemein auf die Struktur der englischen Sprache, und es wäre zu vermuten, dass sich hier nichts für Shakespeare Spezifisches ergibt. Die Methode lässt sich aber prinzipiell auch auf Wort- statt auf Buchstabenketten anwenden und liefert dann ein Verfahren zur Analyse der Komplexität literarischer Texte.

In den 50er und 60er Jahren hat Wilhelm Fucks ein systematisches Forschungsprogramm zu Wort- und Satzstatistiken durchgeführt, wie schon erwähnt (Anm. 2). Man kann beispielsweise die durchschnittliche Anzahl der Buchstaben oder Silben pro Wort oder die Anzahl der Buchstaben oder Wörter in einem Satz zählen. Hier stellt man schon fest, dass sich Prosatexte verschiedener Autoren hinsichtlich solcher Statistiken systematisch unterscheiden. Die quantitative Analyse deckt also gewisse stilistische Unterschiede auf.

Aber auch die Shannon'sche Analyse lässt sich natürlich auf Wortfolgen anwenden. Man geht wieder von der Häufigkeitsverteilung der Wörter in einem Text aus. Hierzu gibt es reichhaltiges Material, auch spezifisch für Shakespeare, und man kann dies auch auf verschiedenen entsprechend gestalteten Webseiten selber ausprobieren. Genauso kann man die Häufigkeit von Wortkombinationen bestimmen. Allerdings wird dies bei dem recht großen Wortschatz einer Sprache oder für unsere Zwecke eines Shakespearetextes schnell rechnerisch sehr aufwändig.

Ein etwas anderer Ansatz untersucht die Häufigkeit des gemeinsamen Vorkommens von Wörtern in einem Text oder Textabschnitt.⁸ Zur Interpretation sind zunächst einmal die von de Saussure entwickelten Grundkonzepte der Sprachwissenschaft hilfreich.⁹ Einerseits können Wörter in einer korrelierten (syntagmatischen) Beziehung zueinander stehen, wie ein Substantiv und ein Adjektiv, und daher typischerweise gemeinsam auftreten. Besonders prägnant ist dies bei idiomatischen oder autorentypischen Wendungen. So spricht Homer immer von der ›kuhägigen

8 Siehe Gerhard Heyer, Uwe Quasthoff und Thomas Wittig: *Text Mining. Wissensrohstoff Text*. Herdecke 2006.

9 Siehe Ferdinand de Saussure: *Grundfragen der allgemeinen Sprachwissenschaft*. Berlin 1967.

Helena«, das heißt, das Adjektiv ›kuhäugig‹ und das Substantiv ›Helena‹ treten sehr häufig gemeinsam auf. Andererseits stehen Synonyme oft in einer komplementären (paradigmatischen) Beziehung, das heißt, weil sie sich gegenseitig ersetzen können, treten sie üblicherweise gerade nicht gemeinsam auf. Allerdings wird letztlich der de Saussure'sche Ansatz dann durch einen quantitativen Ansatz ersetzt, welcher auf relativen Häufigkeiten von Wortkombinationen beruht.

Wortkookkurenzen lassen sich gut in Netzwerken darstellen und visualisieren, und diese Netzwerke lassen sich dann wiederum mit formalen Methoden analysieren.¹⁰

Für unsere Zwecke benötigen wir einige informationstheoretische Konzepte. Wir können diese hier nicht im Detail mathematisch entwickeln, sondern nur deren Zweck und Einsatzmöglichkeiten erläutern.¹¹ Die Shannon'sche Information bemisst, wie stark sich unsere Unsicherheit reduziert, wenn wir ein neues Zeichen sehen, wenn uns vorher nur die erwarteten Häufigkeiten verschiedener möglicher Zeichen bekannt waren, wir aber noch nicht wussten, welches konkrete Zeichen erscheinen würde. Dadurch, dass wir also ein bestimmtes aus einem Ensemble möglicher Zeichen sehen, gewinnen wir Information im Shannon'schen Sinne.

Daher kann auch quantifiziert werden, wieviel Information wir dazugewinnen, wenn wir statt N aufeinanderfolgende $N+1$ Symbole sehen. Dies kann sich abstrakt auf beliebige Zeichenketten beziehen, aber für unsere Zwecke handelt es sich bei den Symbolen um Wörter in einem Text. Mit wachsender Kettenlänge verringert sich natürlich der Informationsgewinn aus der Kenntnis eines weiteren Symbols. Diese Verringerung des zusätzlichen Informationsgewinn als Funktion der Kettenlänge N kann wiederum quantifiziert werden.¹² Wir messen daher in Texten diesen zusätzlichen Informationsgewinn und bestimmen erstens, bei welcher Kettenlänge diese Verringerung am größten wird, nach wievielen Wörtern uns also das nächste Wort die meiste Information liefert, und zweitens, bei welcher Kettenlänge dieser Informationsgewinn von einem Wort zum nächsten am stärksten abfällt. Desweiteren beschreiben geeignete Größen, wieviel Information man über die Fortsetzung eines Textes zu weiteren n Wörtern schon aus der Kenntnis der letzten N Wörter besitzt. Dies sind die von uns verwandten Maße zur Analyse der sprachlichen Komplexität

10 Siehe zum Beispiel Mark Newman: *Networks*. Oxford 2010, oder Anirban Banerjee und Jürgen Jost: »Spectral Plot Properties. Towards a Qualitative Classification of Networks«, *Networks Heterog Media* 3 (2008), S. 395–411.

11 Siehe zum Beispiel Thomas M. Cover und Joy A. Thomas: *Elements of Information Theory*. New York 2006, oder Jürgen Jost: *Dynamical Systems. Examples of Complex Behaviour*. Heidelberg 2005.

12 Vgl. Peter Grassberger: »Toward a Quantitative Theory of Self-Generated Complexity«, *International Journal of Theoretical Physics* 25 (1986), S. 907–938, siehe auch die Darstellung in Jürgen Jost: *Dynamical Systems* (Anm. 11).

von Texten. Es wäre zu spekulieren, und wir wollen belegen, dass bei Shakespeare manche dieser Informationsmaße höhere Werte erreichen als bei durchschnittlichen Autoren.

Die angedeuteten formalen Konzepte sind allgemein genug, um Vergleiche auf unterschiedlichen Skalen zu ermöglichen. Daher lassen sich diese Konzepte nicht nur auf Sprach-, sondern auch auf Handlungsstrukturen in Dramen anwenden. Statt der Abfolge der Wörter in einem Text betrachtet man die Folge der auftretenden und sprechenden Personen. Da die Anzahl der Personen in einem Drama im Vergleich zur Größe des Wortschatzes sehr klein ist, ist dies sogar rechnerisch erheblich einfacher. Die Kookkurenzen (also: wer spricht mit wem?) geben schon Aufschlüsse über die Beziehungen zwischen den handelnden Personen. Die Übergangswahrscheinlichkeiten zwischen Akteuren, das heißt wer nach wem spricht, sollten schon Einiges über den Handlungsablauf verraten. Insbesondere Asymmetrien in diesen Übergangswahrscheinlichkeiten können aufschlussreich sein, wenn also A oft direkt nach B spricht, der umgekehrte Fall aber selten ist. Auch wer in einer Szene das letzte Wort behält, liefert schon wichtige Anhaltspunkte über den nachfolgenden Handlungsablauf. Das Auftreten der verschiedenen Personen werden wir auch durch Verlaufskurven darstellen.

Wie bei den Wortokkurenzen lässt sich auch das gemeinsame Auftreten von Akteuren in einer Szene in einer Netzwerkstruktur darstellen und analysieren. Netzwerke sind inzwischen ein gängiges Werkzeug in den Sozialwissenschaften zur Repräsentation sozialer Interaktionsstrukturen.¹³ Üblicherweise werden dort allerdings nur ungerichtete Netzwerke verwandt, da die Interaktionen typischerweise als symmetrisch angesehen werden; zum Beispiel werden Telefonate oder Emailwechsel zwischen Personen herangezogen. Im Unterschied dazu fassen wir ein Drama als einen Prozess auf, und die Reihenfolge des Auftretens verleiht den Kanten das Netzwerkes dann jeweils eine spezifische Richtung. In diesem Sinne ist unsere literarische Analyse genauer als die soziale Interaktionsanalyse.¹⁴ Allgemeiner sollten durch solche formalen Analysen Anhaltspunkte für die Bewertung der Komplexität einer Handlungsstruktur gewonnen werden.

13 Siehe Matthew Jackson: *Social and Economic Networks*. Princeton 2008.

14 Auch hier lässt sich an eine wichtige Beobachtung von August Wilhelm Schlegel: *Vorlesungen über dramatische Kunst und Literatur* (Anm. 5) S. 140 anknüpfen: »Wenn die Zeichnung seiner Charaktere, jeden einzeln betrachtet, schon unübertrefflich fest und richtig ist, so übertrifft er sich selbst noch in ihrer Zusammenstellung und gegenseitigen Einwirkung«.

IV. Einige formale Punkte, die bei der Analyse beachtet werden müssen

Es ist wichtig, darauf zu achten, dass die formale Analyse nicht bei trivialen Aspekten stecken bleibt. Es ist offensichtlich, dass sich Sprach- und Interaktionsstruktur eines Dramas von derjenigen eines Prosatextes oder eines lyrischen Gedichtes unterscheiden. Daher wird ein formaler Vergleich über verschiedene Textgattungen hinweg zunächst einmal keine wesentlichen Erkenntnisse produzieren. Auch können Dramen in Prosa, Blankversen oder verschiedenen gereimten Versmaßen verfasst werden. Shakespeare setzt diese verschiedenen Formen sehr flexibel ein.¹⁵ Gerade bei den Shakespeareübersetzungen sind dann in dieser Hinsicht die verschiedensten formalen Ansätze zur Anwendung gekommen. Und um dies zu erkennen, ist natürlich kein neues formales Verfahren erforderlich. Auch können verschiedene Theaterbearbeitungen für praktische Zwecke systematische formale Anpassungen vornehmen. All dies zu berücksichtigen ist wichtig für die Interpretation der formalen Befunde.

Ein etwas subtilerer Aspekt ist, dass sich die Länge eines Textes auch als dominanter Effekt bei einer unreflektierten Anwendung von Komplexitätsmaßen herausstellen kann. Um diesen Effekt herauszurechnen, lassen sich unsere formalen Maße daher entweder durch die Anzahl der Wörter oder durch die Anzahl der verschiedenen Wörter normalisieren. Der Anwender kann dann entscheiden, welche Variante sinnvollere Ergebnisse liefert.

V. Einige Ergebnisse

- 1) Wir haben die sprachliche Komplexität der verschiedenen Quartoausgaben der Werke von Shakespeare miteinander verglichen. In den letzten Lebensjahren Shakespeares und in den Jahren nach seinem Tode 1616 sind eine ganze Reihe von Ausgaben erschienen. Es stellt sich heraus, dass diese sich teilweise deutlich in der sprachlichen Komplexität unterscheiden, wobei allerdings die innerhalb des gleichen Jahres verlegten Ausgaben meist sehr ähnlich sind.
- 2) Bei den verschiedenen deutschen Übersetzungen haben wir die sprachliche Komplexität der 3. Szene des 1. Aktes aus Shakespeares *Othello* verglichen. Die ersten Übersetzungen von Wieland und Eschenburg sind noch weniger komplex, aber dann stellt sich ein Komplexitätsniveau ein, welches im weiteren Verlauf gehalten werden kann, bis einige neuere Übersetzungen dieses Niveau wieder verlieren. Einige neuere Theaterbearbeitungen können Abweichungen sowohl nach oben als auch nach unten zeigen.

¹⁵ Ebenda, S. 145 und S. 148 hebt August Wilhelm Schlegel dies als ein von Shakespeare mit besonderer Meisterschaft verwendetes Stilmittel hervor.

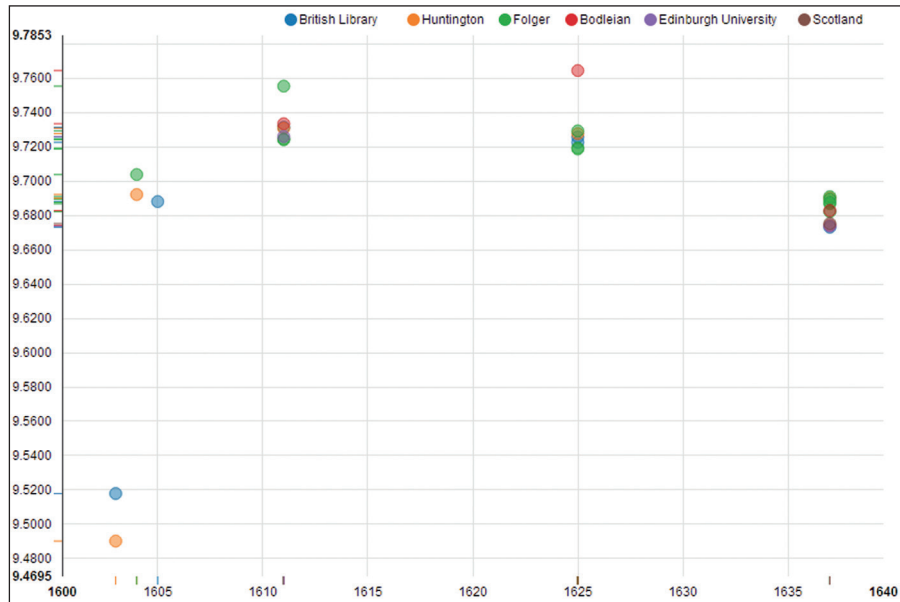
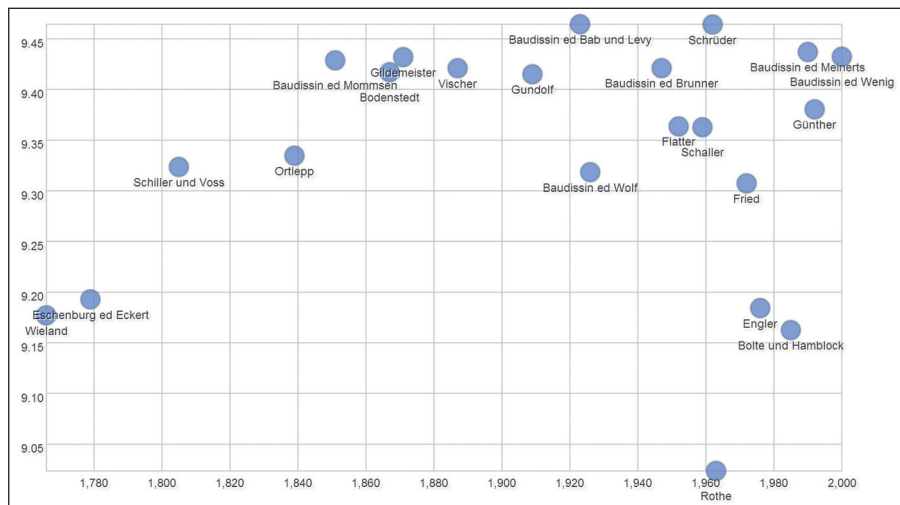


Abb. 1: Sprachliche Komplexität der verschiedenen Quartoausgaben von Shakespeares Werken

Abb. 2: Sprachliche Komplexität deutscher Übersetzungen von *Othello* 1.3

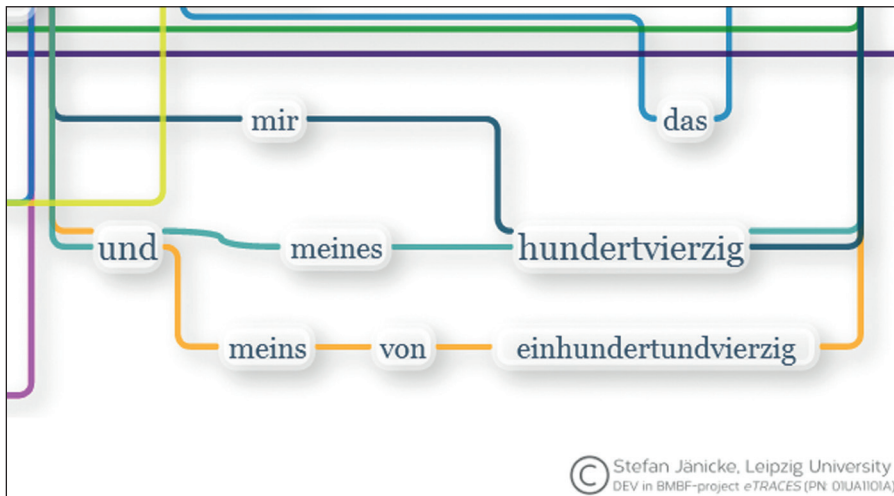
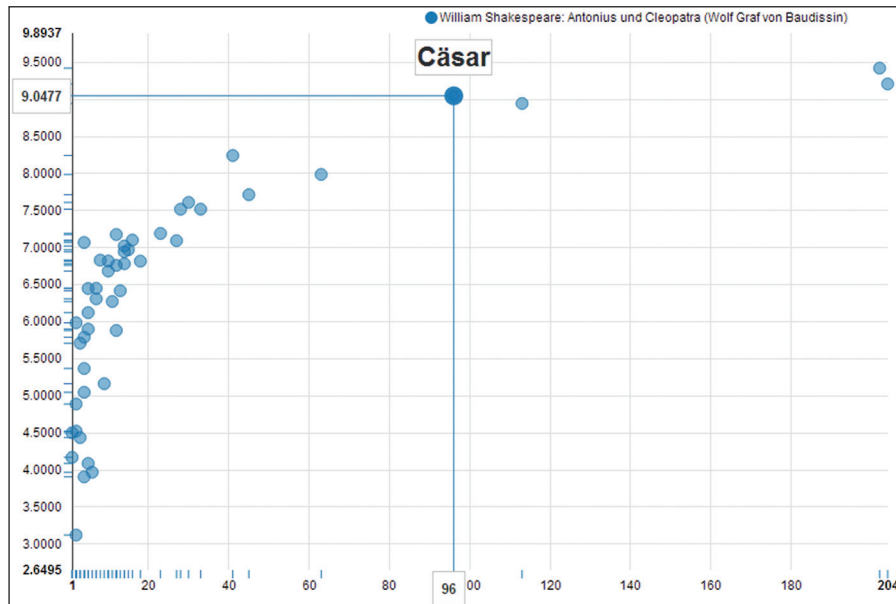
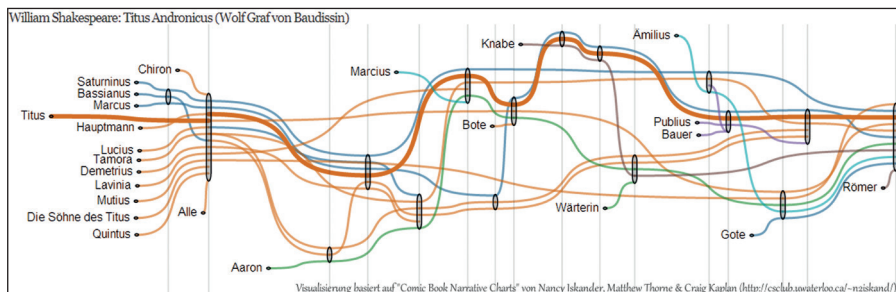


Abb. 3: Wortnutzung in den Übersetzungen am Beispiel von *Orsello* 1.3

- 3) Wir haben dann die Wortnutzung in den Übersetzungen miteinander verglichen, also dargestellt, an welchen Stellen verschiedene Übersetzungen das gleiche Wort benutzen und wo unterschiedliche Wörter verwendet werden. Hier zeigt sich gut das de Saussure'sche Wechselspiel zwischen paradigmatischen Alternativen und syntagmatischen Erfordernissen.¹⁶
- 4) Wir haben die sprachliche Komplexität der verschiedenen Akteure innerhalb eines Stückes miteinander verglichen, um auf automatische Weise die wichtigsten Akteure anhand ihrer sprachlichen Komplexität identifizieren zu können. Hier ist natürlich eine Normalisierung nach der Textlänge der Akteure erforderlich, um zu einem sinnvollen Vergleich gelangen zu können.
- 5) Wir stellen die Verschlungenheit der Handlungsstruktur eines Dramas dadurch dar, dass wir die Szenen des ersten und letzten Auftretens jedes Akteurs durch eine für diesen Akteur farbspezifische Linie miteinander verbinden. Dies visualisiert dann, welche Akteure über welche Zeiträume gemeinsam auftreten. Es zeigen sich insbesondere systematische Unterschiede zwischen Tragödien und Komödien.

¹⁶ Zur Erstellung dieser Abbildung wurde die von Stefan Jänicke, Annette Geßner, Marco Büchler und Gerik Scheuermann: »Visualizations for Text Re-use«. In: *Proceedings of the 5th International Conference on Information Visualization Theory and Applications, IVAPP 2014* (2014), S. 59–70 entwickelte Visualisierungsmethode benutzt.

Abb. 4: Sprachliche Komplexität der Akteure am Beispiel von *Antonius und Cleopatra*Abb. 5: Verschlungenheit der Handlungsstruktur am Beispiel von *Titus Andronicus*

- 6) Wir stellen die Interaktionsstruktur innerhalb eines Dramas durch einen Graphen dar, dessen Knoten die einzelnen Akteure repräsentieren und deren Kantendicke die Anzahl und Länge ihrer Wechselgespräche widerspiegelt. Für das Verständnis der Abbildungen ist es dabei wesentlich, dass die Position der Knoten keinerlei Bedeutung besitzt, sondern dass die wesentlichen Aussagen durch die Stärke und die Struktur der Kantenverbindungen kodiert sind. Die Dicke

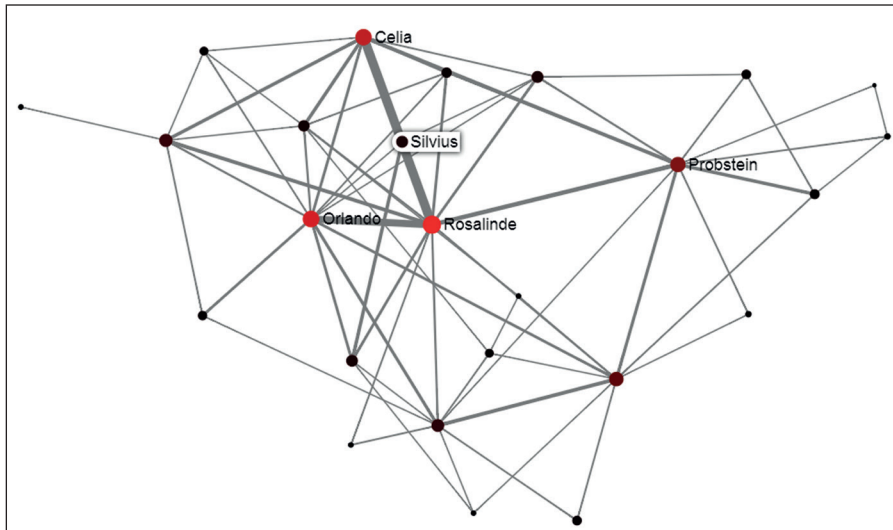


Abb. 6: Interaktionsstruktur innerhalb eines Dramas am Beispiel von *Wie es Euch gefällt*

der Punkte gibt noch die Quantität der Redebeiträge der betreffenden Person an. Durch Setzen oder Variation einer geeigneten Schwelle für die Anzahl der Interaktionen lässt sich dann auch des Zerfallen des Graphen in separate Handlungsstränge erfassen.