



Georg-August-Universität
Göttingen

Philosophische Fakultät

Göttinger philologisches Forum

www.uni-goettingen.de/gpf

Kontakt / Organisation

Berenike Herrmann / Kai Sina / Peer Trilcke
Seminar für Deutsche Philologie
Käte-Hamburger-Weg 3
D-37073 Göttingen
jb.herrmann@phil.uni-goettingen.de
kai.sina@phil.uni-goettingen.de
trilcke@phil.uni-goettingen.de

Thesenpapier zum Vortrag am 13. Dezember 2012

Dr. Christof Schöch (Würzburg)

Stylometrische Experimente. Autorschaft und Gattungszugehörigkeit im französischen Theater der Klassik

Quantitative Verfahren der Analyse von Sammlungen literarischer Texte haben in den letzten Jahren bedeutende Fortschritte gemacht. Dies ist sicherlich der vermehrten Verfügbarkeit geeigneter digitaler Texte sowie der Existenz zunehmend nutzerfreundlicher und leistungsfähiger Werkzeuge zu verdanken, die Verfahren wie die Stylometrie oder das Topic Modeling wesentlich erleichtern. Immer mehr Literaturwissenschaftler experimentieren mit solchen Verfahren und die Menge an Erfahrungswerten und Einsatzmöglichkeiten steigt. Gleichzeitig sind die methodischen Schwierigkeiten und Unwägbarkeiten immer noch enorm, gerade für Literaturen in anderen Sprachen als dem Englischen.

Der Vortrag wird von der Erprobung stylometrischer Klassifikationsverfahren im Bereich des französischen Theaters des siebzehnten Jahrhunderts berichten. Hierfür wurden Jan Rybickys und Maciej Eders stylometrischen Skripte für die Statistikumgebung R eingesetzt. Ausgangspunkt ist die mittlerweile berühmte Corneille-Molière-Kontroverse, die einen faszinierenden Testfall für stylometrische Verfahren darstellt. Im Zentrum der Kontroverse steht die Frage, ob Corneille einige oder sehr viele der traditionellen Molière zugeschriebenen Werke verfasst hat. Einerseits erfordert eine solche Frage die Lösung sehr konkreter Probleme, wie das des relativen Einflusses von Autorschaft und Gattungszugehörigkeit bei stylometrischen Klassifikationssaufgaben. Von einigen Versuchen, solche Fragen methodisch in den Griff zu bekommen, wird hauptsächlich die Rede sein. Der stylometrische Ansatz wirft andererseits aber auch sehr allgemeine Fragen der Verlässlichkeit, Nachvollziehbarkeit und Relevanz statistischer Verfahren in den Philologien auf, die ebenfalls diskutiert werden sollen.

Zugespielt formuliert lassen sich drei Thesen ins Spiel bringen:

- (1) Die Stylometrie ist im Kern ein Verfahren, das weniger der Stilistik, als der Literaturgeschichte nahesteht: es geht ihr nicht um die Beschreibung von Texteigenschaften an sich, sondern um die Klassifikation von Texten auf der Grundlage von Texteigenschaften.
- (2) Die Corneille/Molière-Kontroverse ist unentscheidbar, solange es nicht gelingt, das Verhältnis der Signale von Autorschaft, Gattung und Form zueinander zu verstehen und Methoden zu erproben oder zu entwickeln, um diese Signale zu unterscheiden. Dies ist eine der aktuellen Herausforderungen der Stylometrie.

- (3) Die Verfahren der Digitalen Geisteswissenschaften müssen für alle an einem Projekt beteiligten Personen im Detail nachvollziehbar sein; dies erfordert nicht, dass wir alle vollwertige Programmierer, Statistiker und Philologen zugleich werden; vielmehr gilt es, in kleinen heterogenen Teams die eigenen disziplinären Stärken und Spezialisierungen einzubringen; Bedingung ist allerdings, dass alle Beteiligten mit den jeweils anderen kommunizieren lernen.

Lektürehinweise

Burrows, John. "‘Delta’: a Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17.3, 2002: 267-287.

Brunet, Étienne. "Où l'on mesure la distance entre les distances." *Texto!*, 04/2004. http://www.revue-texto.net/Inedits/Brunet/Brunet_Distance.html.

Labbé, Cyril, and Dominique Labbé. "Inter-textual Distance and Authorship Attribution. Corneille and Molière." *Journal of Quantitative Linguistics* 8/3, 2001, 213–231.

Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London/New York: Verso, 2005.

Rybicki, Jan, and Maciej Eder. "Deeper Delta Across Genres and Languages: Do We Really Need the Most Frequent Words?" *Literary and Linguistic Computing* 26/3, 2011, 315–321.

Ressourcen

Eder, Maciej & Jan Rybicki, "stylo" & "classify" scripts for R, v. 0.4.6, 2012, <https://sites.google.com/site/computationalstylistics/>.

Fièvre, Paul, dir.: *Théâtre classique*, 2012, <http://www.theatre-classique.fr/>.

"Quantitative Text Analysis", in: *DARIAH-Bibliographie*, DARIAH-DE, version 2, 04/20012. <http://kurzlink.de/text-analysis>.

