

Für eine computergestützte literarische Gattungsstilistik

Christof Schöch und Steffen Pielström (Computerphilologie, Univ. Würzburg)

DHd-Konferenz, Passau, 25.-28. März 2014

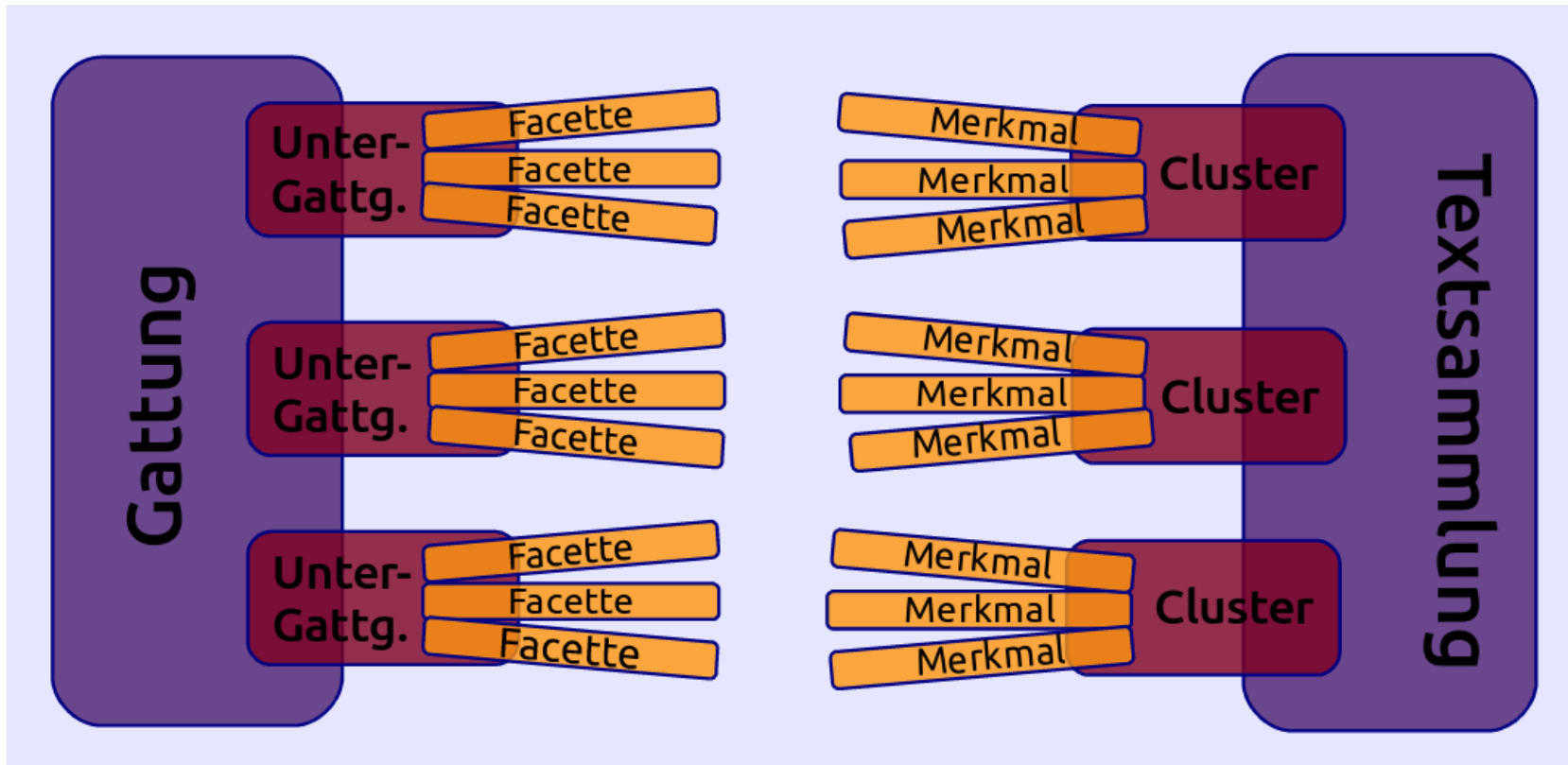
Gliederung

- Einleitung
- I. Computergestützte literarische Gattungsstilistik
- II. Konzeptuelle Ebene: Verknüpfung von Gattungstheorie und Stilistik
- III. Methodische Ebene: Erweiterung der Principal Component Analysis
- Fazit

I. Forschungsfragen der computergestützten literarischen Gattungsstilistik

- **Literaturwissenschaftliche Fragestellungen**
 - Beziehung zwischen Stil und Gattung?
 - Sprachliche Merkmale von Gattungen?
 - Gattungsstile, Epochenstile und Autorenstile?
- **Methodische Fragestellungen**
 - Trennung von Autoren- Epochen und Gattungssignal?
 - Welche computergestützten Verfahren sind am besten geeignet?
 - Wie können Robustheit, Transparenz, Interpretierbarkeit, Validierbarkeit erweitert werden?
- **Weiterführende Fragestellungen**
 - Verhältnis von Interpretation und Algorithmus?

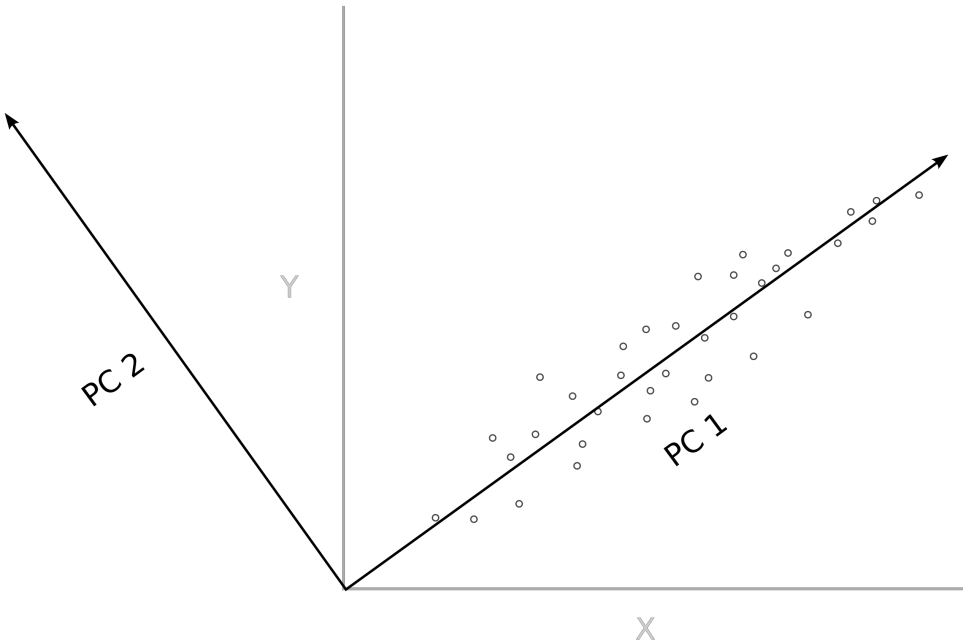
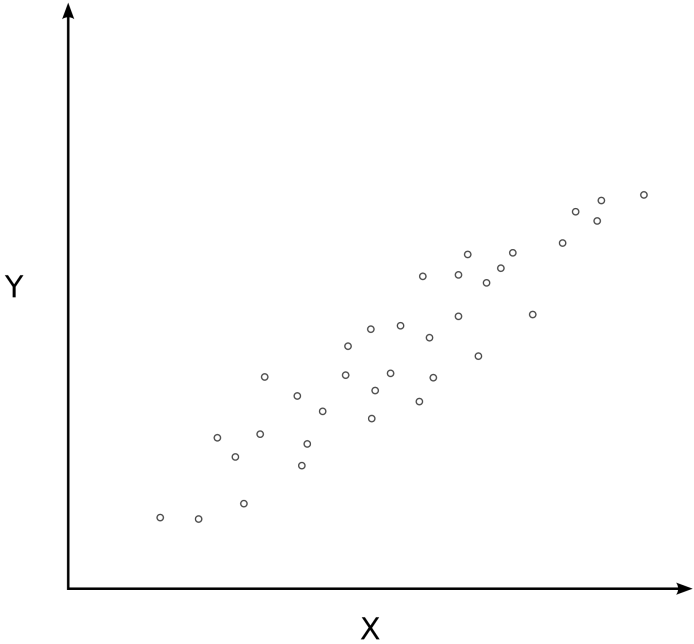
II. Gattungstheorie und computergestützte Stilistik



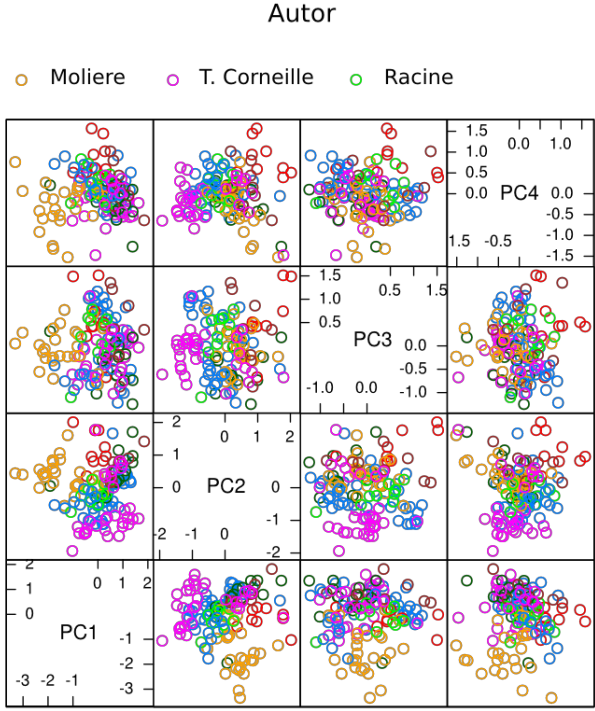
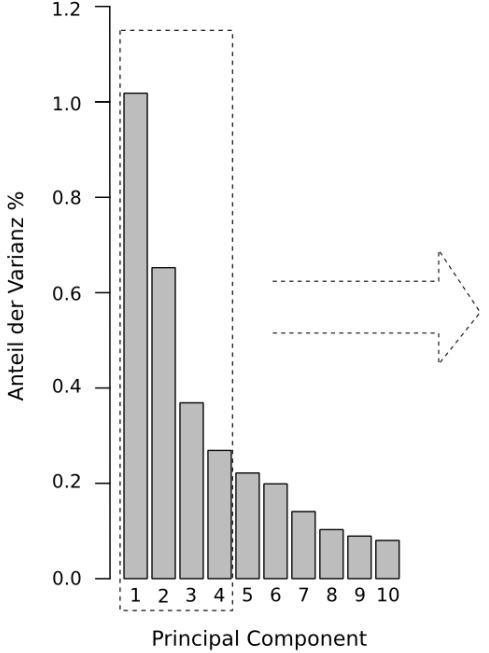
III. Methodische Erweiterung der Principal Component Analysis

	A	B	C	D	E
1		CBoyer_Federic	CBoyer_Tyridate	Moliere_Avare	Moliere_Tartuffe
2	de	3.2182054939	2.5497847147	3.7960122699	2.9577620522
3	et	2.3811000186	2.4623250807	2.8326303681	3.201017249
4	vous	2.2818875178	2.2604951561	3.369440184	3.6930561698
5	que	1.8106281391	1.8366523143	2.9572469325	1.9073418841
6	je	1.7114156384	1.9779332616	2.6984279141	2.1118973905
7	l	1.7052148571	1.7693756728	1.2845092025	1.4705882353
8	le	1.599801575	2.1057588805	1.4474693252	1.5037593985
9	à	1.6556086067	1.7020990312	1.7062883436	1.9073418841
10	un	2.3252929869	1.4666307858	1.0017254601	1.4484741265
11	il	1.1781484467	1.5069967707	1.7254601227	1.6530296329
12	est	1.2401562597	1.2177072121	1.7158742331	1.5922158337

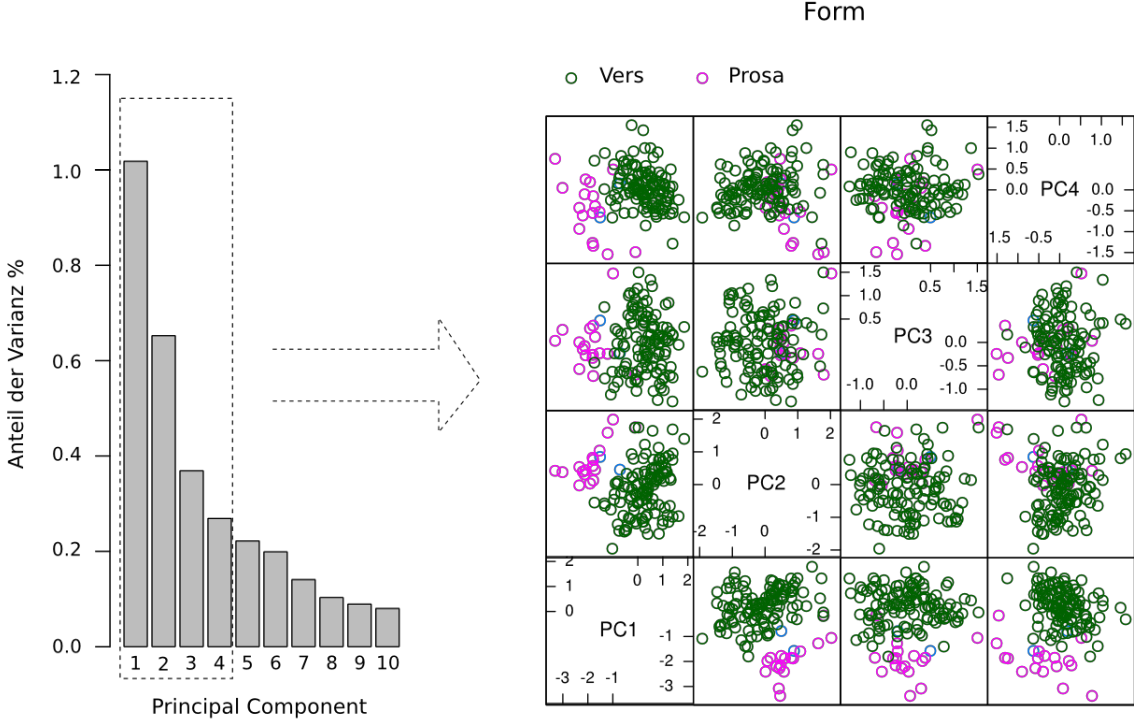
III. Methodische Erweiterung der Principal Component Analysis



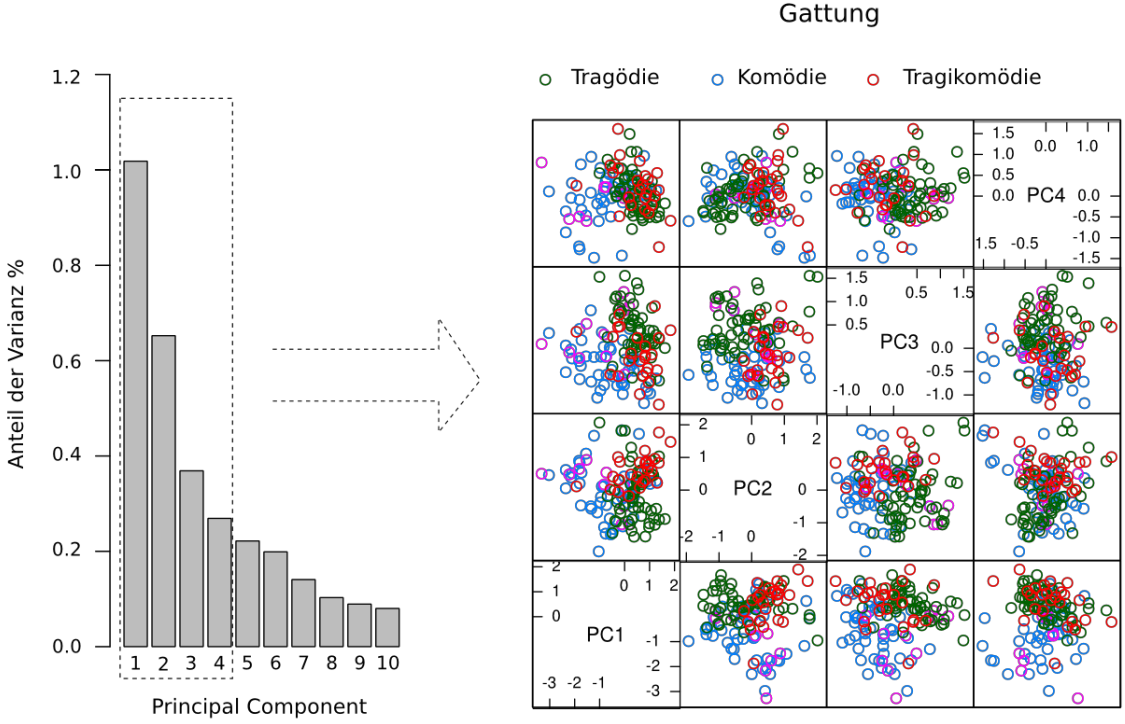
Erster Schritt: Scree Plot und Scatterplot Matrix



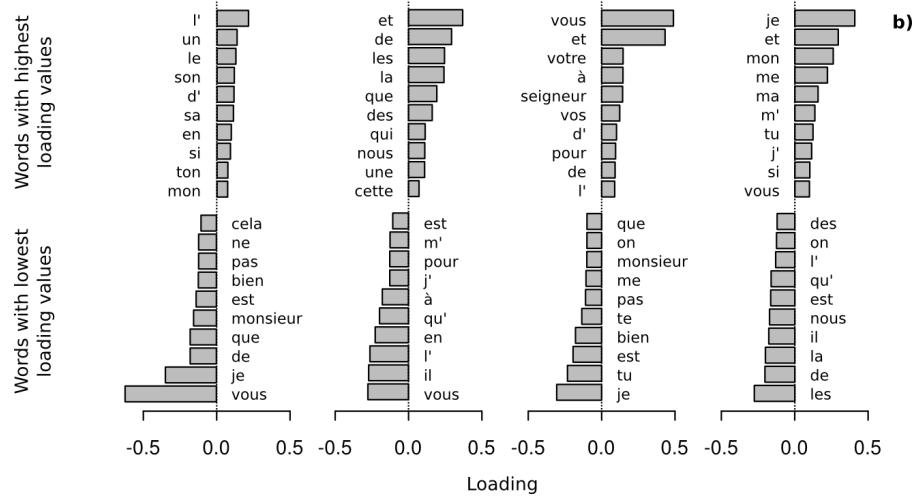
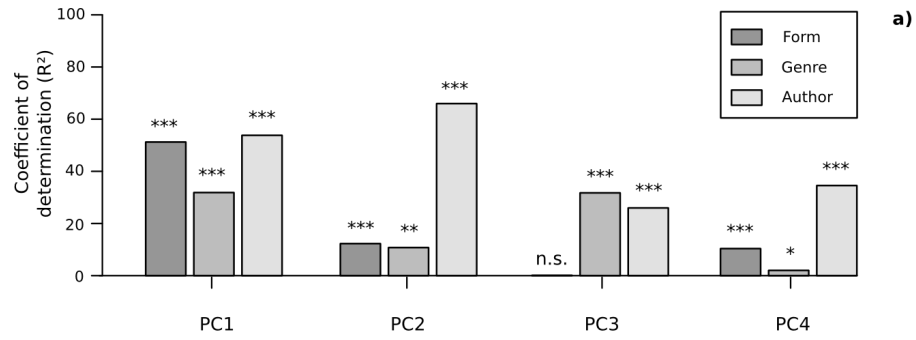
Erster Schritt: Scree Plot und Scatterplot Matrix



Erster Schritt: Scree Plot und Scatterplot Matrix



Zweiter Schritt: PCA-plus: Blick in die Komponenten



Fazit

- Enge Verbindung von Literaturwissenschaft und Statistik
- Signifikanz (statistisch): zuverlässige und robuste Ergebnisse
- Signifikanz (literaturwiss.): interpretierbare, bedeutungsvolle Ergebnisse
- Blick für stilistische Details, angewandt auf große Textmengen

Bibliografie

- Allison, Sarah, Ryan Heuser, Matthew L. Jockers, Franco Moretti, and Michael Witmore (2011). *Quantitative Formalism: An Experiment*. Stanford: Stanford Literary Lab.
- Biber, Douglas (1992). "The multidimensional approach to linguistic analyses of genre variation", in: *Computers in the Humanities*, 26.56, 331-347.
- Burrows, John (2002). "Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship". *Literary and Linguistic Computing* 17.3, 267-287.
- Combe, Dominique (2002). "La stylistique des genres", in: *Langue française* 135, 33-49.
- Hoffmann, Michael (2009). "Mikro und makrostilistische Einheiten im Überblick", in: *Rhetorik und Stilistik, Ein internationales Handbuch historischer und systematischer Forschung, Band 2*, hg. von Ulla Fix, Andreas Gardt & Joachim Knape. Band 2, Berlin: de Gruyter, 1529-45.
- Jackson, Edward (2005). *A User's Guide to Principal Components*. New York: Wiley.
- Jockers, Matthew (2013). *Macroanalysis. Digital Methods and Literary History*. Chicago: Univ. of Illinois Press.
- Juola, Patrick (2006). "Authorship Attribution." *Foundations and Trends in Information Retrieval* 1.3, 233-334.
- Karlgren, Jussi & Douglas Cutting (1994). "Recognizing text genres with simple metrics using discriminant analysis". In *Proceedings of COLING '94, Vol. 2*, 1071--1075.
- Kessler, Brett, Geoffrey Numberg, and Hinrich Schütze (1998). "Automatic Detection of Text Genre." In *Proceedings of ACL 1998*, 3238. doi:10.3115/976909.979-622.
- Kestemont, Mike, Kim Luyckx, Walter Daelemans, and Thomas Crombez (2012). "CrossGenre Authorship Verification Using Unmasking." *English Studies* 93.3, 340-356.
- Moretti, Franco (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.
- Sandig, Barbara (2006). *Textstilistik des Deutschen*. 2. Auflage. Berlin: de Gruyter.
- Schaeffer, JeanMarie (1989). *Qu'est-ce qu'un genre littéraire?* Paris: Seuil, 1989.
- Schöch, Christof (2013). "Finetuning Our Stylometric Tools: Investigating Authorship and Genre in French Classical Theater", *Digital Humanities Conference 2013*, <http://dh2013.unl.edu/abstracts/ab270.html>.
- Stamatatos, Efsthios, Nikos Fakotakis, and George Kokkinakis (2000). "Automatic Text Categorization in Terms of Genre and Author." *Computational Linguistics* 26/4, 471-497.