

STILOMETRIE IN DER RECHTSLINGUISTIK

Nutzung korpuslinguistischer Verfahren für die Analyse deutschsprachiger Urteile

Anna-Maria Auer / Pascale Berteloot / Bettina Mielke /
Christine Schikora / Thomas Schmidt / Christian Wolff

B.A., Institut für Information und Medien, Sprache und Kultur, Lehrstuhl für Medieninformatik
Universität Regensburg, 93040 Regensburg, DE
anna-maria.auer@stud.uni-regensburg.de, <http://mi.ur.de>

Dr. iur. h.c., Dozentin an den Universitäten Köln und Trier (ehem. Abteilungsleiterin beim Amt für Veröffentlichungen der EU)
145 rue Albert Unden, 2652 Luxembourg, LU
bertepl@pt.lu

Dr. iur., Vorsitzende Richterin am Landgericht Regensburg, Lehrbeauftragte an der Universität Regensburg
Kumpfmühler Straße 4, 93047 Regensburg, DE
bettina.mielke@lg-r.bayern.de

B.A., Institut für Information und Medien, Sprache und Kultur, Lehrstuhl für Medieninformatik
Universität Regensburg, 93040 Regensburg, DE
christine.schikora@stud.uni-regensburg.de, <http://mi.ur.de>

M. Sc., wiss. Mitarbeiter, Institut für Information und Medien, Sprache und Kultur, Lehrstuhl für Medieninformatik
Universität Regensburg, 93040 Regensburg, DE
thomas.schmidt@ur.de, <http://mi.ur.de>

Professor, Institut für Information und Medien, Sprache und Kultur, Lehrstuhl für Medieninformatik
Universität Regensburg, 93040 Regensburg, DE
christian.wolff@ur.de, <http://mi.ur.de>

Schlagworte: *Korpuslinguistik, Rechtssprache, Stilometrie, Digital Humanities, Europäischer Gerichtshof*

Abstract: *Mit aktuellen korpuslinguistischen Verfahren analysieren wir eine Sammlung von Entscheidungen des Europäischen Gerichtshofs in deren deutscher Fassung. In Fortsetzung zum Beitrag Berteloot/Mielke/Wolff 2018 nehmen wir stilometrische Analyseverfahren in den Blick, die in den digitalen Geisteswissenschaften intensiv genutzt werden, um z.B. das Problem der Zuordnung von Autorenschaft zu untersuchen. Daran angelehnte Verfahren konnten im US-amerikanischen Bereich bereits erfolgreich auf die Analyse von Urteilen des Supreme Court übertragen werden. Wir stellen für unser Beispielkorpus ausgewählte Ergebnisse solcher stilometrischer Analysen vor und diskutieren deren Aussagekraft.*

1. Einführung

Aktuelle Studien aus den USA zeigen, wie sich Urteile des obersten amerikanischen Gerichts (*Supreme Court*) mit Hilfe von Verfahren aus dem Bereich des Text Mining analysieren lassen (CARLSON/LIVERMORE/ROCKMORE 2015; LIVERMORE/RIDDELL/ROCKMORE 2017; ROSENTHAL/YOON 2010). Unter anderem ist dabei untersucht worden, wie sich Autorenprofile von Richtern und ihrem juristischen Unterstützungspersonal (*law clerks*), die sich über eine Analyse des Gebrauchs von Funktionswörtern ermitteln lassen (ROSENTHAL/YOON 2010), im Lauf der Zeit geändert haben. Auch vergleichsweise neue Verfahren des Text Mining wie etwa *topic modeling* (BLEI/NG/JORDAN 2003) oder *word embeddings* (PENNINGTON/SOCHER/MANNING 2014) sind dabei erprobt worden.

In unserem Beitrag nehmen wir dies zum Anlass, eine rechtslinguistische Studie zur deutschen Rechtssprache auf europäischer Ebene weiter fortzusetzen (vgl. BERTELOOT/MIELKE/WOLFF 2018). Wir stellen dazu ein Korpus aus Entscheidungen des Europäischen Gerichtshofs in deren deutscher Fassung zusammen und wenden darauf in den digitalen Geisteswissenschaften bereits vielfach eingesetzte Verfahren der Stilometrie an (EDER/RYBICKI/KESTEMONT 2016; EDER 2015; HOLMES/KARDOS 2003).

Während in den Geisteswissenschaften versucht wird, mittels solcher Methoden die Autorenschaft zu klären, so z.B. auch die oben genannte amerikanische Studie, rücken aus juristischer oder rechtslinguistischer Perspektive weitere Probleme in den Mittelpunkt, etwa Fragen nach der Homogenität der Urteilssprache oder auch den Vorlieben für Modalpartikel oder andere Funktionswörter in den verschiedenen Varietäten des Deutschen.

2. Stilometrie

Stilometrie hat seinen Ursprung in dem Forschungsgebiet der Autorenschaftszuweisung und bezeichnet den Einsatz computergestützter Methoden zur Quantifizierung und statistischen Analyse des literarischen Stils in textuellen Datensätzen (HOLMES/KARDOS 2003; STAMATATOS 2009). Ziele sind dabei meist die eindeutige Identifikation und Charakterisierung von Autoren sowie Ähnlichkeitsanalysen, auch im Kontext von Plagiatidentifikation (EL BOUANANI/KASSOU 2014). Stilometrie hat eine lange Tradition, die bis zu (damals manuell durchgeführten) Autorenschaftsanalysen bezüglich Shakespeares Dramen zurückreicht (MENDENHALL 1897). Neben der Klärung von literaturwissenschaftlichen Fragestellungen zu literarischen Texten wird Stilometrie heutzutage auch erfolgreich im Bereich von E-Mails (GOODMAN 2007), Texten aus sozialen Medien (VOSOUGHI/ZHOU/ROY 2015) oder Programmcode eingesetzt (CALISKAN-ISLAM et al. 2015).

Das methodische Vorgehen bei stilometrischen Studien besteht aus mehreren aufeinanderfolgenden Schritten: Vorverarbeitung und Aufbereitung des Textkorpus, Merkmalsextraktion, statistische Analyse und Vergleich sowie Visualisierung der Ergebnisse. Von zentraler Bedeutung ist dabei stets der zweite Schritt, also die korrekte Identifikation stilometrischer Merkmale zum Vergleich der textuellen Untersuchungsgegenstände. STAMATATOS 2009 identifiziert folgende Merkmalskategorien: *Lexikalische* (z.B. Worthäufigkeiten, Wort-N-Gramme), *zeichenbasierte* (z.B. Zeichen-N-Gramme), *syntaktische* (z.B. Wortarten, Satzstrukturen), *semantische* (z.B. semantische Abhängigkeiten) und *anwendungsspezifische* Merkmale (z.B. bei der Analyse von Webseiten die Häufigkeit von HTML-Tags). Die Mehrzahl stilometrischer Studien verwendet lexikalische Merkmale, wengleich sich der Einsatz zeichenbasierter Merkmale vor allem im Bereich moderner Textarten wie E-Mails als erfolgreich herausgestellt hat (STAMATATOS 2009). Die Kombination verschiedener Merkmale aus allen Kategorien ist ebenfalls möglich (IQBAL et al. 2008).

Bezüglich der Repräsentation des Stils wird unterschieden zwischen profilbasierten und instanzenbasierten Ansätzen (STAMATATOS 2009). Für profilbasierte Ansätze wird eine größere Menge an Textdokumenten eines Autors zur weiteren Analyse zusammengestellt. Diese Textmenge bestimmt das «Profil» des Autors und neue Texte werden bezüglich der Ähnlichkeit zu diesem Profil untersucht. Bei instanzenbasierten Ansätzen wird jedes Dokument isoliert betrachtet und die Ähnlichkeit zwischen den einzelnen Dokumenten analysiert (STAMATATOS 2009). Zur weiteren mathematischen Analyse und Visualisierung des Stils werden je nach Anwendungsfall unterschiedliche Techniken zur Vorhersage der Autorenschaft und zur Analyse von Ähnlichkeiten verwendet wie z.B. Methoden des maschinellen Lernens, Clustering-Algorithmen und probabilistische Modelle (EL BOUANANI/KASSOU 2014).

Die häufigste Methodik in aktuellen Studien ist ein instanzenbasierter Ansatz unter Nutzung eines Vektorraummodells und von Distanzmaßen zur Identifikation der intertextuellen Distanz von Dokumenten. Auch in der vorliegenden Arbeit wird diese Methodik eingesetzt. Die einzelnen Textdokumente werden dazu in einem multidimensionalen Vektorraum mit den einzelnen Merkmalen als Dimensionen und den Ausprägungen (Stärke des Vorkommens des jeweiligen Merkmals in einem Dokument) als Vektoren repräsentiert. Im Fall von Worthäufigkeiten kann man beispielsweise die 100 häufigsten Wörter einer Dokumentmenge als Dimensio-

nen verwenden und die Häufigkeiten als Vektoren. Daneben können auch bekannte Klassifikationsalgorithmen wie *support vector machines* (SVM) und neuronale Netzwerke verwendet werden. Bei Ähnlichkeitsbasierten Modellen werden mathematische Distanzmaße im Vektorraum genutzt, um für alle repräsentierten Texte ihre intertextuelle Ähnlichkeit festzustellen und zu visualisieren. Einige klassische Distanzmaße sind die Manhattan-, die euklidische oder die Cosinus-Distanz. Im Bereich der Stilometrie hat sich vor allem das von BURROWS 2002 entwickelte Ähnlichkeitsmaß «Delta», häufig auch «Burrows Delta» genannt, durchgesetzt. Dieses ist vor allem bei Texten, die länger als 1500 Wörter sind, nachweislich effektiv (STAMATATOS 2009).

Im Bereich der Visualisierung der Ergebnisse gibt es unterschiedliche Möglichkeiten, oft auch abhängig von der verwendeten statistischen Analyse. Am bekanntesten sind auf Cluster-Analyse basierende Dendrogramme, «*consensus trees*», Repräsentationen auf 2D-Achsen, Heatmaps und vor allem bei größeren Dokumentmengen Netzwerkanalysen (BENJAMIN et al. 2014; EDER 2017).

3. Korpusaufbau

Der Korpusaufbau erfolgte durch Auswahl von Urteilen des Europäischen Gerichtshofs aus der Datenbank EURLEX zu einem der klassischen Bereiche des EU-Rechts, nämlich Freizügigkeit und freier Dienstleistungsverkehr. Diese Themen haben teilweise auch als Ausgangspunkt für Clustering-Versuche gedient (vergleiche unten Kap. 6). Es handelt sich um ca. 90 Urteile aus dem Jahr 2018, von denen jedoch nur ein Teil für die tatsächliche Analyse genutzt wurde.

4. Forschungsfragen

Da uns bislang keine ähnlichen Arbeiten im Bereich stilometrischer Analysen von Urteilstexten im deutschsprachigen Bereich bekannt sind, ist der Forschungsansatz dieser Arbeit explorativ. Auf Basis verfügbarer Metadaten zu den Urteilstexten des Korpus werden Teilkorpora erstellt, die nach Einschätzung der Autoren für stilometrische Analysen geeignet sind. Dazu werden gruppenbasierte Vergleiche mit einer überschaubaren Anzahl an Dokumenten und maximal drei unterschiedlichen Gruppen von Dokumenten durchgeführt. Jedes Dokument wird gemäß Ausprägung eines Metadatum einer Gruppe/Kategorie zugeordnet. In der stilometrischen Analyse wird sodann untersucht, ob die Dokumente einzelner Gruppen stilistisch zueinander «clustern», ähnlich sind oder sich von Dokumenten einer anderen Gruppe unterscheiden.

Folgende drei Analysen wurden ausgewählt und durchgeführt. Zum einen wird über das Metadatum des Landes, aus dem das Ursprungsurteil stammt, ein Subkorpus bestehend aus den Urteilen, deren Ursprungsurteil aus Deutschland, und den Urteilen, die ihren Ursprung in Österreich haben. Vergangene Arbeiten (MIELKE/WOLFF 2016, BERTELOOT/MIELKE/WOLFF 2018) hatten sprachliche Unterschiede zwischen der österreichischen und der deutschen Rechtssprache sowie mögliche Unterschiede in deutschsprachigen Urteilen des Europäischen Gerichtshofs zum Gegenstand. Hier wird untersucht, ob sich Unterschiede bezüglich stilometrischer Analysen zeigen.

Als zweite Analyse wird die Kategorie des dem Urteil zugeordneten Generalanwalts («*advocate general*») beispielhaft untersucht. Als konkreter Anwendungsfall werden die Urteile in den Rechtssachen, die Generalwältin Juliane Kokott und Generalanwalt Paolo Mengozzi zugeordnet sind, untersucht und hinsichtlich stilometrischer Besonderheiten analysiert.

Als letztes wird eine stilometrische Analyse bezüglich des angegebenen Themas («*subject matter*») durchgeführt. Diesbezüglich wurde das Korpus auf Dokumente gefiltert, die gemäß angegebenem Metadatum nur einem Thema zuzuweisen sind; die Dokumente, die den drei häufigsten Einzelthemen zugeordnet sind, wurden miteinander verglichen. Bei den Themen handelt es sich um Niederlassungsrecht («*right of establishment*»), freier Dienstleistungsverkehr («*freedom to provide services*») und Sozialversicherung für Arbeitsmigranten («*social security for migrant workers*»).

Es handelt sich in allen drei Fällen um kleinere überschaubare Subkorpora, die adäquat für die explorative erste Untersuchung der Methodik der Stilometrie in diesem Anwendungsgebiet sind.

5. Methodik

Zur methodischen Umsetzung der stilometrischen Analysen orientieren wir uns an den gängigen Standard-Methoden des Forschungsfeldes, da ähnliche Studien und damit methodische Vorarbeiten und Erfahrungen fehlen. Dementsprechend haben wir uns für einen instanzbasierten Ansatz entschieden, bei dem überschaubare Mengen an Dokumenten untereinander verglichen werden. Zum mathematischen Vergleich verwenden wir ein Vektorraummodell mit Burrows Delta als Distanzmaß, das sich als Standard in der Stilometrie etabliert hat und insbesondere für Texte geeignet ist, die länger als 1500 Wörter sind (STAMATATOS 2009), was für alle Dokumente des Korpus zutrifft. Diese Methodik ist vor allem dann geeignet, wenn grundlegende Ähnlichkeiten untersucht werden sollen und keine Autorenschaftszuweisungen durchgeführt werden (müssen).

Bezüglich der Merkmale empfiehlt STAMATATOS 2009 die Verwendung der Worthäufigkeiten der häufigsten Wörter eines Korpus. Erste Erfolge können dabei bereits bei einem Set der 100 häufigsten Wörter festgestellt werden. Bei längeren Dokumenten werden deutlich mehr Wörter verwendet, z.B. die 1000 häufigsten (STAMATATOS 2009). Je nach Anwendungsfall werden statt einzelner Wörter N -gramme verwendet, also die Folge von jeweils N Wörtern im laufenden Text, wobei häufig verwendete Werte für N zwei (Bigramme) oder drei (Trigramme) sind. Da wir für unser Korpus vor allem einen Einfluss komplexerer Wortformulierungen annehmen, nutzen wir für die Analysen Trigramme, also die Häufigkeiten von Abfolgen von je drei Wörtern anstatt einfacher Worthäufigkeiten. Insgesamt haben wir für unser Korpus und unsere Fragestellungen die besten Ergebnisse bei Verwendung der 1000 häufigsten Trigramme festgestellt. In einem Analysefall erhöhen wir die Menge der Trigramme, da eine höhere Zahl von Dokumenten betrachtet wird (siehe unten Kap. 6). Ferner haben wir die Methodik des «Culling» angewandt. Dabei werden N -Gramme entfernt, die bezüglich eines Grenzwerts in zu wenigen Dokumenten vorkommen. Wir verwenden einen Culling-Wert von 10%, d.h. N -Gramme müssen in mindestens 10% der Dokumente einer Analyse vorkommen, um betrachtet zu werden. Auf diese Weise werden N -Gramme entfernt, die (sehr) häufig genutzt werden, jedoch nur in einer sehr kleinen Anzahl an Dokumenten. Im Bereich der Stilometrie ist der Einsatz von Culling-Werten eine Standard-Methodik (EDER/RYBICKI/KESTEMONT 2016; GLADWIN/LAVIN/LOOK 2015).

Bezüglich der technischen Umsetzung verwenden wir die Programmiersprache R mit dem *stylo*-Paket. Das *stylo*-Paket ist ein R-Paket für stilometrische Analysen (EDER/RYBICKI/KESTEMONT 2016). Es bietet Funktionen für alle notwendigen Analysen. Alle Auswertungen und Visualisierungen wurden mit diesem Paket umgesetzt.

Zur Visualisierung der Ergebnisse verwenden wir Dendrogramme (Baumdiagramme) basierend auf hierarchischer Clusteranalyse (EDER 2017). Dabei wird basierend auf dem Distanzmaß und den Merkmalen eine Hierarchie von Clustern über einen «bottom-up»-Ansatz gebaut. Die sich paarweise ähnlichsten Dokumente werden zu einem Cluster verbunden und je nach mathematischer Nähe auf nachfolgenden Ebenen werden weitere einzelne oder bereits gruppierte Dokumente hinzugefügt. In den nachfolgenden Visualisierungen werden die Gruppenzugehörigkeiten einzelner Dokumente über unterschiedliche Grauwerte gekennzeichnet, um die Interpretation zu vereinfachen.

6. Ergebnisse und Interpretation

Nachfolgend werden für die drei Auswahlkriterien einzelne Analysebeispiele für die Subkorpora vorgestellt. Wiedergegeben ist dabei jeweils ein Dendrogramm einer spezifischen Analyse, dessen Blätter jeweils für einzelne Dokumente stehen, die durch die jeweilige Merkmalsausprägung (z.B. Austria, Germany) und ihre Dokumentnummer charakterisiert sind.

6.1. Urteile in Vorlageverfahren aus Deutschland und Österreich

In der ersten Analyse wird ein Subkorpus, bestehend aus allen Urteilen in Vorlageverfahren aus Österreich und Deutschland untersucht. Es besteht aus 17 Urteilstexten, davon neun nach deutschen und acht nach österreichischen Vorlagen. Die stilometrische Analyse nach der oben beschriebenen Methodik wird über ein Dendrogramm basierend auf hierarchischer Cluster-Analyse in Abbildung 1 illustriert. Beide Gruppen (Deutschland und Österreich) werden farblich differenziert. Die Nummern am Ende des Kategoriennamens dienen als Identifikationsnummern für die einzelnen Dokumente (Urteile) über alle gruppenbasierte Vergleiche hinweg:

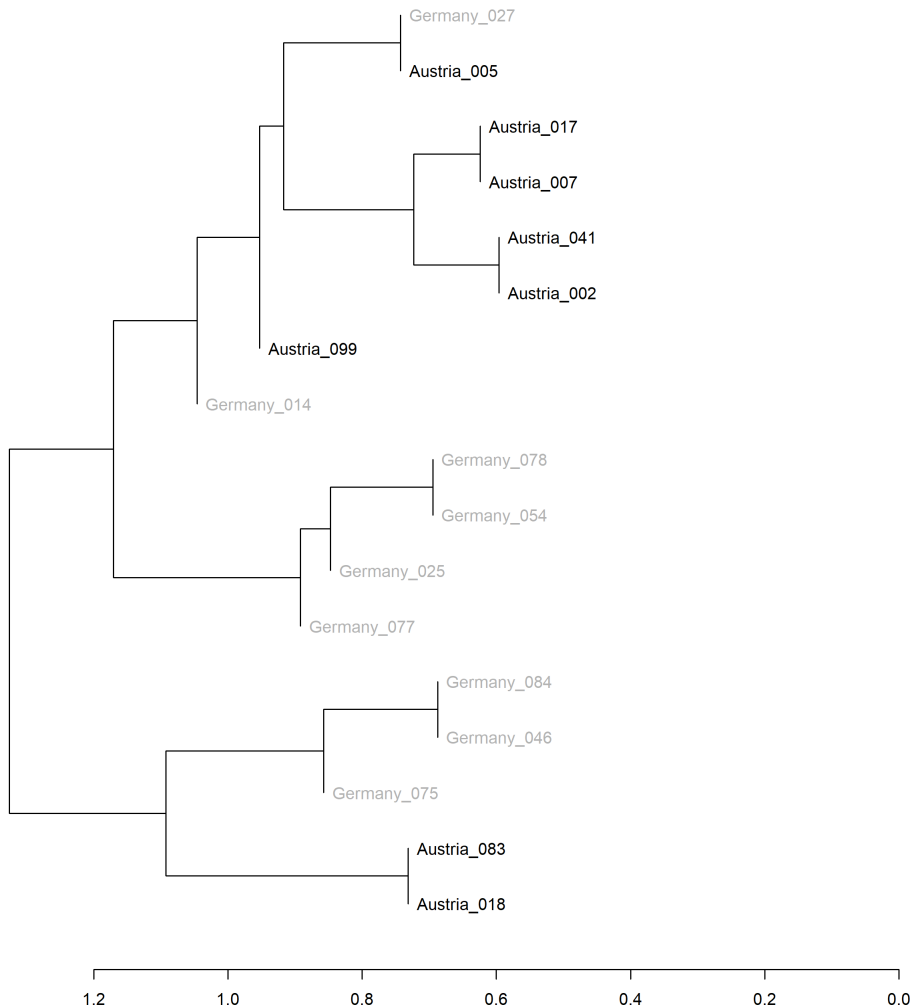


Abbildung 1: Clustering von Teilkorpora auf der Basis der Herkunft des Vorlageverfahrens

Es werden drei größere Sub-Cluster deutlich. Das oberste ist stark dominiert von österreichischen Dokumenten mit zwei deutschen Dokumenten. Das mittlere besteht nur aus Dokumenten mit Ursprung in Deutschland, während das dritte gemischt ist. Dennoch ist deutlich, dass sich auf der ersten Ebene bis auf eine Ausnahme stets gleiche Kategorien am stärksten ähneln.

6.2. Generalanwalt: Juliane Kokott und Paolo Mengozzi

Das Subkorpus bezüglich der Analyse der Urteile in Rechtssachen, die entweder Generalanwältin Kokott oder Generalanwalt Mengozzi zugeordnet sind, besteht aus 21 Urteilen (elf Fälle mit einer Zuordnung zu Generalanwalt Mengozzi und zehn mit einer Zuordnung zu Generalanwältin Kokott). Abbildung 2 illustriert die Ergebnisse der stilometrischen Analyse. Da die Erfahrung lehrt, dass in vielen Fällen der Gerichtshof in seinen Urteilen den Generalanwälten zumindest im Ergebnis folgt (ARREBOLA/MAURICION/JIMÉNEZ PORTILLA 2016), erscheint dieses Analysekriterium durchaus erprobenswert.

Insgesamt liegt ein sehr homogenes kategorienkonformes Ergebnis vor. Auffällig ist die Differenzierung der drei unteren Dokumente (Kokott_031-033). Weiterführende, hier nicht dokumentierte, Analysen mit anderen Generalanwälten und Kategorien zeigen, dass diese drei Dokumente immer miteinander «clustern» und sich von anderen Dokumentmengen stark abgrenzen. Der übrige Cluster lässt sich deutlich in zwei Teile mit jeweils einem Ausnahmedokument teilen, einmal der Cluster Mengozzi mit dem Dokument Kokott_011 und der Kokott-Cluster mit dem Dokument Mengozzi_009.

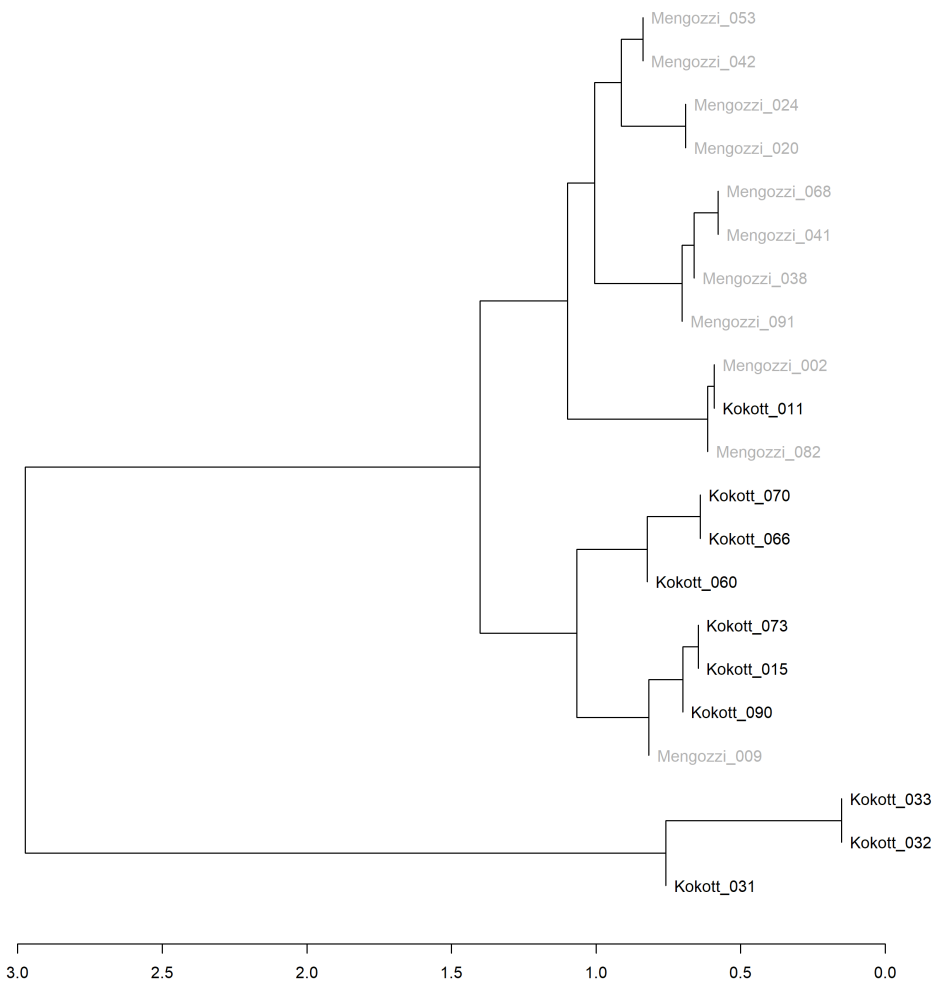


Abbildung 2: Cluster der personenbezogenen Subkorpora (Generalanwälte Kokott / Mengozzi)

6.3. Thematik

Das dritte Subkorpus, das nach der Kategorie Thematik gebildet wurde, besteht aus 23 Dokumenten und ist damit das größte Subkorpus. Es wird bezüglich dreier Themen differenziert: Niederlassungsrecht (neun Dokumente), freier Dienstleistungsverkehr (sechs Dokumente) und Sozialversicherung für Arbeitsmigranten (acht Dokumente; im folgenden Sozialversicherung genannt). Aufgrund der Größe haben wir die Zahl der häufigsten Trigramme auf 2000 erhöht. Abbildung 3 illustriert das Ergebnis.

Das Ergebnis ist deutlich heterogener als in den anderen Fällen. Während kleinere Subcluster bezüglich der Themen Niederlassungsrecht und Sozialversicherung zu erkennen sind, verteilt sich das Thema Dienstleistungsverkehr deutlich stärker. Das Ergebnis zeigt auch, dass die stilometrische Analyse mit zunehmender Dokumentanzahl schwieriger zu interpretieren ist.

7. Diskussion und Ausblick

Der explorative Charakter einer solchen Untersuchung zeigt sich nicht zuletzt darin, dass durch Anpassung der numerischen Parameter des Analyseverfahrens eine Vielzahl von Einzelergebnissen produziert wird, aus denen dann «bessere» Baumdiagramme intellektuell ausgewählt werden, ohne den inhaltlichen Zusammenhang zwischen Parametersetzung und Ergebnisqualität unmittelbar erklären zu können. Eine inhaltliche Analyse über die zugrundeliegenden Daten, aus denen das Dendrogramm erzeugt wird, also konkret die Zuordnungstabellen zwischen den nach Häufigkeit sortierten betrachteten Trigrammen und den Werten der Zuordnungsstärke zum jeweiligen Dokument ist aufwendig. Es lässt sich sagen, dass Analysen wie oben präsentiert noch kein für sich selbst stehendes Ergebnis liefern, sondern erste Hinweise für weitergehende Analysen und intellektuelle Vergleiche zwischen den einzelnen Dokumenten. Die drei Analysebeispiele, die hier vorgestellt werden, erfüllen den Zweck der explorativen Studie: Sie zeigen die grundsätzliche Anwendbarkeit der stilometrischen Analyseverfahren, können aber keinesfalls eine tiefere Auseinandersetzung mit den Möglichkeiten dieser Analyseform für bestimmte juristische Fachtextsorten wie Urteile, die einer bestimmten Kommunikationssituation entstammen und für die in der Regel ein vergleichsweise präzise beschreibbarer Aufbereitungsprozess existiert, ersetzen.

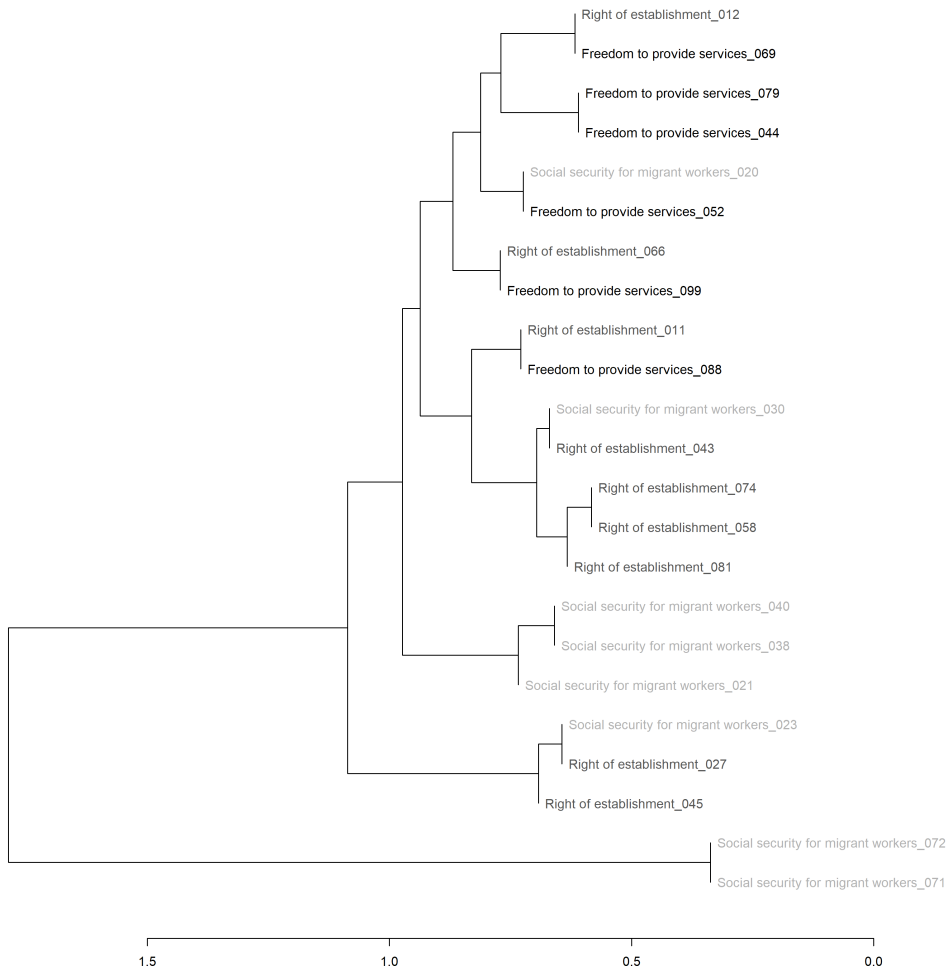


Abbildung 3: Ergebnisse der thematischen Clusteranalyse für drei Themenschwerpunkte

8. Literatur

ARREBOLA, CARLOS/MAURICIO, ANA JULIA/JIMÉNEZ PORTILLA, HÉCTOR, An Econometric Analysis of the Influence of the Advocate General on the Court of Justice of the European Union, *Cambridge Journal of Comparative and International Law*, Jg. 5, Heft , Forthcoming; University of Cambridge Faculty of Law Research Paper No. 3/2016. Online: <https://ssrn.com/abstract=2714259>.

BENJAMIN, VICTOR et al., Evaluating Text Visualization for Authorship Analysis. *Security Informatics*, 2014, Jg. 3, Heft 1, S. 10.

BERTELOOT, PASCALE/MIELKE, BETTINA/WOLFF, CHRISTIAN, Deutsches, österreichisches, europäisches Deutsch? Deutschsprachige Fassungen von Urteilen des europäischen Gerichtshofs im Vergleich. In: Erich Schweighofer/Franz Kummer/Ahti Saarenpää/Burghardt Schafer (Hrsg.), *Datenschutz / LegalTech. Data Pro-*

- tection / LegalTech. Tagungsband des 21. Internationalen Rechtsinformatik Symposions IRIS 2018, Editions Weblaw, Bern 2018, S. 319–324.
- BLEI, DAVID M/NG, ANDREW Y/JORDAN, MICHAEL I, Latent Dirichlet Allocation, *Journal of Machine Learning research*, 2003, Jg. 3, S. 993–1022.
- BURROWS, JOHN, «Delta»: a Measure of Stylistic Difference and a Guide to likely Authorship. *Literary and linguistic computing*, 2002, Jg. 17, Heft 3, S. 267–287.
- CALISKAN-ISLAM, AYLIN et al., De-anonymizing Programmers via Code Stylometry. In: 24th USENIX Security Symposium (USENIX Security), Washington, DC, 2015, S. 255–270.
- CARLSON, KEITH/LIVERMORE, MICHAEL A/ROCKMORE, DANIEL, A Quantitative Analysis of Writing Style on the US Supreme Court, *Washington University Law Review*, 2015, Jg. 93, S. 1461–1510.
- EDER, MACIEJ, *Rolling Stylometry*, *Digital Scholarship in the Humanities*, 2015, Jg. 31, S. 457–69.
- EDER, MACIEJ, *Visualization in Stylometry: Cluster Analysis Using Networks*. *Digital Scholarship in the Humanities*, 2017, 32. Jg., Heft 1, S. 50–64.
- EDER, MACIEJ/KESTEMONT, MIKE/RYBICKI, JAN, *Stylometry with R: a Suite of Tools*. In: *Digital Humanities 2013: Conference Abstracts*. Lincoln, NE: University of Nebraska, Lincoln, 2013. S. 487–489.
- EDER, MACIEJ/RYBICKI, JAN/KESTEMONT, MIKE, *Stylometry with R: A Package for Computational Text Analysis*, *R Journal*, 2016, Jg. 8, Heft 1, S. 107–121.
- EL BOUANANI, SARA/KASSOU, ISMAIL, *Authorship Analysis Studies: A Survey*. *International Journal of Computer Applications*, 2014, Jg. 86, S. 22–29.
- GLADWIN, ALEXANDER/LAVIN, MATTHEW J./LOOK, DANIEL M., *Stylometry and Collaborative Authorship: Eddy, Lovecraft, and «The Loved Dead»*. *Digital Scholarship in the Humanities*, 2017, Jg. 32, Heft 1, S. 123–140.
- GOODMAN, ROBERT et al., *The Use of Stylometry for Email Author Identification: a Feasibility Study*. In: *Proc. Student/Faculty Research Day, CSIS, Pace University, White Plains, NY, 2007*, S. 1–7.
- HOLMES, DAVID I./KARDOS, JUDIT, *Who was the Author? An Introduction to Stylometry*. *Chance*, 2003, 16. Jg., Heft 2, S. 5–8.
- IQBAL, FARKHUND et al., *A Novel Approach of Mining Write-prints for Authorship Attribution in E-mail Forensics*. *Digital Investigation*, 2008, 5. Jg., S. S42–S51.
- LIVERMORE, MICHAEL A/RIDDELL, ALLEN B/ROCKMORE, DANIEL N, *The Supreme Court and the Judicial Genre*, *Arizona Law Review*, 2017, Jg. 59, S. 837–901.
- MENDENHALL, THOMAS CORWIN, *The Characteristic Curves of Composition*, *Science*, 1887, 9. Jg., Heft 214, S. 237–249.
- MIELKE, BETTINA/WOLFF, CHRISTIAN, *Österreichische und Deutsche Gerichtsentscheidungen im Sprachvergleich*. In: *Erich Schweighofer/Franz Kummer/Walter Hötendorfer/Georg Borges (Hrsg.), Netzwerke. Proceedings 19. Internationales Rechtsinformatik-Symposium Salzburg (IRIS 2016), Österreichische Computer-Gesellschaft (ÖCG), Wien 2016*, S. 129–138.
- PENNINGTON, JEFFREY/SOCHER, RICHARD/MANNING, CHRISTOPHER, *GloVe: Global Vectors for Word Representation*. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014, S. 1532–1543.
- ROSENTHAL, JEFFREY S/YOON, ALBERT H., *Judicial Ghostwriting: Authorship on the Supreme Court*, *Cornell Law Review*, 2010, Jg. 96, S. 1307–1344.

STAMATATOS, EFSTATHIOS, A Survey of Modern Authorship Attribution Methods. *Journal of the American Society for information Science and Technology*, 2009, 60. Jg., Heft 3, S. 538–556.

VOSOUGHI, SOROUSH/ZHOU, HELEN/ROY, DEB, Digital Stylometry: Linking Profiles Across Social Networks. In: *International Conference on Social Informatics*. Springer, Cham, CH, 2015, S. 164–177.