

Benedikt Bönninghoff, Steffen Hessler, Dorothea Kolossa, Kerstin Kucharczik, Robert M. Nickel, Karin Pittner

# Autorschaftsanalyse

## Verstellungsstrategien und Möglichkeiten der automatisierten Erkennung

Der Austausch von Informationen findet zunehmend über soziale Medien und Online-Nachrichtendienste statt. Daraus kann ein hohes Schadenspotential erwachsen, wenn

es gelingt, für kriminelle Absichten eine andere Identität vorzutäuschen oder im Schutze der Anonymität Fake News bzw. Hate Speech zu verbreiten. Zum Schutz vor solchen Angriffen untersucht die Forensische Linguistik Textsammlungen hinsichtlich der Urheberschaft und möglicher biographischer Aussagen. Doch der Umfang der über das Internet veröffentlichten Daten verlangt automatisierte Verfahren, die die Analyse unterstützen.

### 1 Einführung

Ziel unseres Projektes ist es, automatisierte Verfahren zu entwickeln, die zur Analyse der Autorschaft von Texten mit unklarer Herkunft sowie zur Aufdeckung von Täuschungsversuchen oder



#### Benedikt Bönninghoff

wissenschaftlicher Mitarbeiter der Arbeitsgruppe Kognitive Signalverarbeitung an der Ruhr-Universität Bochum und Doktorand im Forschungskolleg SecHuman. Forschungsschwerpunkte: Künstliche Intelligenz, Computerlinguistik, Audioverarbeitung.

E-Mail: Benedikt.Boenninghoff@rub.de



#### Steffen Hessler

wissenschaftlicher Mitarbeiter am Lehrstuhl für Germanistische Linguistik an der Ruhr-Universität Bochum und Doktorand im Forschungskolleg SecHuman. Forschungsschwerpunkte: Forensische Linguistik, Varietätenlinguistik, Korpuslinguistik, Dialektologie, Linguistic Awareness im Bereich IT-Security.

E-Mail: Steffen.Hessler@rub.de



#### Dorothea Kolossa

Professorin für Kognitive Signalverarbeitung an der Ruhr-Universität Bochum, Mitglied im Research Department Horst-Görtz-Institut für IT-Sicherheit und im Exzellenzcluster CaSa an der Ruhr-Universität Bochum sowie Principal Investigator im

Forschungskolleg SecHuman.  
E-Mail: Dorothea.Kolossa@rub.de



#### Kerstin Kucharczik

Oberrätin für Germanistische Linguistik an der Ruhr-Universität Bochum sowie Principal Investigator im Forschungskolleg SecHuman.

E-Mail: Kerstin.Kucharczik@rub.de



#### Robert M. Nickel

Professor für Elektrotechnik an der Bucknell University, Lewisburg, Pennsylvania (USA) und VIP Fellow der Ruhr-Universität Research School PLUS. Forschungsschwerpunkte: Künstliche Intelligenz und Digitale Signalverarbeitung.

E-Mail: rmn009@bucknell.edu



#### Karin Pittner

Professorin für Germanistische Linguistik an der Ruhr-Universität Bochum sowie Principal Investigator im Forschungskolleg SecHuman.

E-Mail: Karin.Pittner@rub.de

falschen Identitäten eingesetzt werden können. Dabei verfolgen wir einen interdisziplinären Ansatz, bei dem Methoden und Ergebnisse der linguistischen Sprachanalyse mit Verfahren aus dem Gebiet der künstlichen Intelligenz verknüpft werden.

Ein Ausgangspunkt unseres Projekts war die Beobachtung, dass bei der medialen Stilisierung von sprachlichen Varietäten oder Regiolekten, wie z. B. dem Ruhrdeutschen von Comedians, bestimmte, als besonders charakteristisch empfundene Merkmale herausgegriffen und sehr häufig verwendet werden, andererseits aber nicht dazu passende Merkmale auftreten. Auch beim Internet-Betrug oder bei Erpresserbriefen können Verschleierungsstrategien und Vortäuschung falscher Identitäten daran erkannt werden, dass z. B. augenfällige Merkmale von „Ausländerdeutsch“ bzw. „Foreigner talk“ (vgl. [6, 8]) auftreten und daneben Merkmale, die auf eine/n Muttersprachler/in schließen lassen. Untersucht werden soll, ob die Vortäuschung falscher Identitäten im Internet (z. B. in Bezug auf Herkunft oder Alter) durch Häufungen charakteristischer Merkmale sowie durch nicht dazu passende Merkmale zu erkennen ist.

### Beispiele für Verstellungsstrategien:

Imitation von Migrantendeutsch in inkriminierten Texten: *Wenn ischo dich seh, ischo mach disch Messer.*

Stilisierung der gesprochenen Sprache im Ruhrgebiet: *„Ich geh mit im Zimmer, da läufte Film, nä, Sie – nich eine Minute konntich datt sehen, da wa einer ne Hand am Grab am raus am Ziehen – nänänä, ersparensset mir, ich sach Heidi, Kint, mach datt aus, datt ertrachich nich.“* (Beleg aus Heidenreich, Elke (1984): „Darf’s ein bißchen mehr sein?“ Rowohlt Verlag, S. 76; zum Gebrauch von am-Progressiv-Formen (im Beleg unterstrichen) in der gesprochenen Sprache im Ruhrgebiet vgl. [12])

Will man Verstellungsstrategien mit maschinellem Lernen erkennen, so besteht die Aufgabe darin, Algorithmen zu entwickeln, die die von uns vermuteten strukturellen Merkmale mit hoher Zuverlässigkeit in verschiedenen Textsorten auffinden und Dokumente entsprechend klassifizieren können. In diesem Beitrag stellen wir den Zusammenhang von verschiedenen Verstellungsstrategien und das automatische Erkennungssystem Ad-Hominem vor. Wir haben es uns zur Aufgabe gemacht, Forschungsergebnisse vorzulegen, die in Bereichen der Verbrechensbekämpfung, der Abwehr von Social Engineering wie Phishing bzw. CEO-Fraud und zur Steigerung der IT-Security-Awareness an Universitäten, in Behörden und Unternehmen eingesetzt werden können.

Unser Praxispartner ist die Abteilung für Autorenerkennung des Bundeskriminalamts in Wiesbaden, die ständig mit Fragen der Autorschaft inkriminierter Texte wie Erpresser- und Bekenner schreiben oder Texten aus dem Bereich des politischen Extremismus befasst ist, die heute vorwiegend über das Internet verbreitet werden.

## 2 Forensische Linguistik

In vielen schriftsprachlichen Bereichen des Internets, wie Texten in sozialen Medien, Online-Rezensionen oder anonymen E-Mails, haben die Leser/innen keine gesicherten Informationen über die Verfasser/innen. Im mündlichen Bereich erhalten Hörer/innen viele Hinweise auf außersprachliche Merkmale (Metadaten), wie Herkunft, Alter oder Geschlecht der Sprecher/innen.

Viele Hörer/innen haben z. B. eine ungefähre Vorstellung, aus welchem Bereich des deutschen Sprachraums Sprecher/innen kommen, da u. U. eine bestimmte Dialektregion erkannt wird. Solche Informationen sind bei der Rezeption von Texten nicht bzw. nur erschwert zugänglich. Die Autorenerkennung aus dem Fachbereich der Forensischen Linguistik (vgl. für eine Übersicht u. a. [2, 13] und für den deutschsprachigen Raum [3, 7]) befasst sich mit der Bewertung fraglicher Texte und kann dabei Aufschlüsse über die Autorschaft geben.

### 2.1 Textanalyse vs. Textvergleich

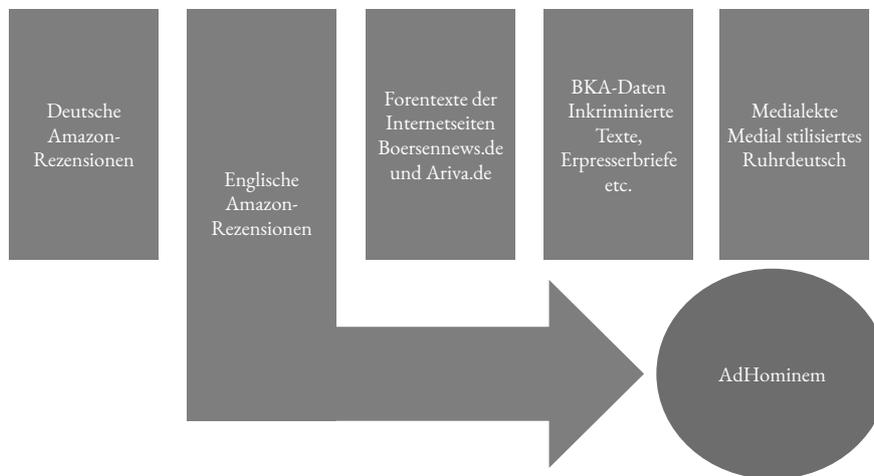
Bei der Autorenerkennung werden zwei Hauptbereiche unterschieden: die Textanalyse und der Textvergleich (vgl. insbesondere [3, 5, 7]; Einblicke in die privatwirtschaftliche Perspektive bietet [4]). Bei der Textanalyse ist das Ziel, anhand eines oder mehrerer Texte Anhaltspunkte über Alter, Geschlecht, Herkunft, Bildungsgrad etc. einer/eines Autorin/Autors zu erhalten. Dahingegen werden beim Textvergleich einer oder mehrere Texte unbekannter Herkunft einem oder mehreren Texten bekannter Autor/inn/en gegenübergestellt. Hier wird anhand der Analyse sprachlicher Merkmale der Grad der Übereinstimmung bestimmt. Die Autorenerkennung des Bundeskriminalamts unterscheidet beispielsweise sechs Grade der Übereinstimmung, wobei die Grade von „non liquet“ (unklar) bis „mit an Sicherheit grenzender Wahrscheinlichkeit“ reichen (vgl. auch [3] S. 76). Obwohl Autor/inn/en über ein recht individuelles Merkmal-Set verfügen können, kann es keine hundertprozentige Sicherheit über die Autorschaft eines Textes geben. Die Vorstellung eines linguistischen Fingerabdrucks ist daher irreführend. Vielmehr kann man von einem Fingerzeig sprechen (zur Individualstil-Debatte vgl. [3] S. 53 ff. und [7] S. 121 ff.). Bei sogenannten Sammlungsrecherchen werden inkriminierten Schreiben aus archivierten Strafverfahren des Bundeskriminalamtes zum Vergleich herangezogen.

Bei der Textanalyse werden sprachliche Merkmale wie Fehler, Abweichungen von der Standardsprache und sonstige Auffälligkeiten in verschiedenen Bereichen der Textproduktion wie Grammatik, Rechtschreibung, Zeichensetzung, Stilistik, Formatierung und Layout analysiert. Beim Textvergleich können fragliche Schreiben mit sogenannten Vergleichsschreiben, bei denen die Autorschaft sicher ist, verglichen werden. Ziel des Vergleichs ist es, eine Aussage über eine mögliche gleiche Autorschaft zwischen zwei oder mehreren Texten zu treffen.

### 2.2 Datenlage und Erstellung von Korpora

Es wurden verschiedene Korpora erstellt und die Texte auf Vorkommen linguistischer Merkmale analysiert. Zu den Korpora gehören Amazon-Rezensionen in deutscher und englischer Sprache. Der Fokus der Analyse deutschsprachiger Rezensionen liegt auf dem Vorkommen und dem Nachweis von regiolektalen Merkmalen, die Hinweise darauf geben, in welcher deutschsprachigen Region die Autor/inn/en sprachlich sozialisiert wurden. Ein weiteres Korpus besteht aus Texten von Online-Foren, in denen ein Autor scheinbar unter mehreren Nicknames Beiträge verfasst. Weiterhin wurden im Rahmen der Kooperation mit dem BKA verschiedene inkriminierte Texte analysiert. Herzstück der Analyse ist hier ein Fall, der aus 49 Texten (u. a. Erpresserschriften) besteht und dabei ein breites Feld von Merkmalen und verschiedene Verschleierungs- und Imitationsstrategien aufweist,

Abb. 1 | Datensätze und Korpora



wie etwa die Imitation von Migrantendeutsch („Kanak Sprach“). Dabei ähneln die stilisierten Merkmale meist nicht der Sprache von Deutschlernenden, sondern der Sprache von Muttersprachlern, wenn sie mit Ausländern reden, dem sogenannten „Foreigner talk“. Als Vertreter der Medialekte wurde ein großes Korpus aus Texten erstellt, die dem medial stilisierten Ruhrdeutsch zugeordnet werden können. Diese Texte bieten die größte quantitative und qualitative Variation an Merkmalen und Verstellungsstrategien. Die englischsprachigen Rezensionen dienen als Grundlage für das Training des automatischen Systems AdHominem, das wir in 3.4 näher vorstellen.

### 2.3 Verstellungsstrategien und Analysebereiche sprachlicher Merkmale

Da es im schriftsprachlichen Bereich keine offensichtlichen Hinweise auf die Metadaten von Autor/inn/en gibt, ist es hier auch leichter möglich, seinen eigenen Sprachgebrauch zu verstellen (vgl. [1]) und den Sprachgebrauch von Gruppen bzw. Einzel-

personen zu imitieren bzw. zu stilisieren, um falsche Identitäten vorzutäuschen. Bei Texten, die in krimineller Absicht verfasst werden, wie beispielsweise Erpresserbriefen, (Spear)-Phishing-Mails oder CEO-Frauds (zu Betrugsschreiben bzw. „Fraud Cases“ vgl. [14]) ergibt sich hieraus ein großes Schadenspotential. Sprachliche Verstellungsstrategien werden ebenfalls bei Fake-Rezensionen, Hate Speech und Fake News sowie beim Cybermobbing, beim „Trollen“ und „Flamen“ in sozialen Netzwerken und Internetforen eingesetzt.

Im Rahmen unserer Forschung werden Methoden im Bereich der Forensischen Linguistik entwickelt, mit denen Verstellungsstrategien erforscht und aufgedeckt werden können. Bei der Nachahmung anderer Personen oder der Imitation vermeintlich typischer Vertreter/innen bestimmter Gruppen werden atypische bzw. überstilisierte Merkmale verwendet, die nicht zu den sprachlichen Registern passen, die die nachgeahmte/n Person/en in bestimmten Situationen verwenden würde/n. Es gibt Hinweise auf Verstellungsstrategien, wie die Verwendung spezifischer Orthographie- und Grammatikfehler, eine auffällige Stilistik wie die Verwendung bestimmter Merkmale von Dia-, Regio-, und Soziolekten, Fachsprachen sowie Registern.

Bei der schriftsprachlichen Verstellung unterscheiden wir drei verschiedene Bereiche (vgl. Abb. 2).

**Verschleierung:** Bei der Verschleierung der eigenen Identität verfolgen Autor/inn/en das Ziel, dass die Adressat/inn/en ihres Textes keine Rückschlüsse auf ihre Autorschaft ziehen können. Beispielsweise verwendet ein User in einem Forum einen Hauptaccount und mehrere Fake-Accounts. In Diskussionen unterstützen die Fake-Accounts die Aussagen des Hauptaccounts, damit diese für andere User/innen mehr Gewicht bzw. Gültigkeit erhalten. Oder der/die User/in trollt bzw. flamet andere Nutzer/innen,

ohne befürchten zu müssen, dass sein/ihr eigener Account gesperrt wird. Die Strategie der Verschleierung der eigenen Identität wird auch in u. U. strafrechtlich relevanten Bereichen wie Hate Speech oder Fake-Rezensionen eingesetzt.

**Imitation:** Im Falle der Imitation anderer Identitäten tritt eine weitere Dimension der Verstellung hinzu. Hier versucht ein/e Autor/in nicht nur, die eigene Identität nicht preiszugeben, sondern imitiert die Schreibweise anderer Gruppen oder Einzelpersonen. In Erpresserschreiben wird die Schriftsprache beispielsweise so angepasst, dass ein/e muttersprachliche/r Autor/in versucht, wie ein/e Nicht-Muttersprachler/in zu schreiben oder umgekehrt. Diese Strategie hängt stark zusammen mit der Imi-

Abb. 2 | Dimensionen schriftsprachlicher Verstellungen

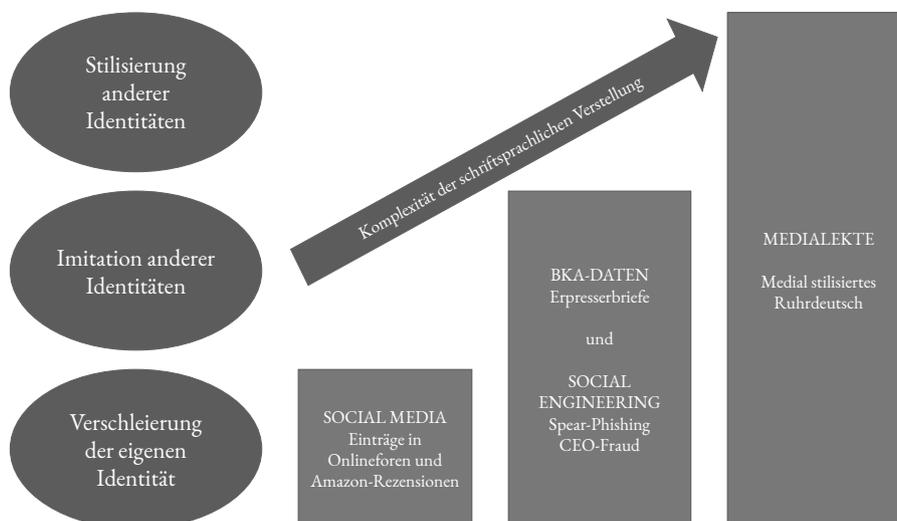
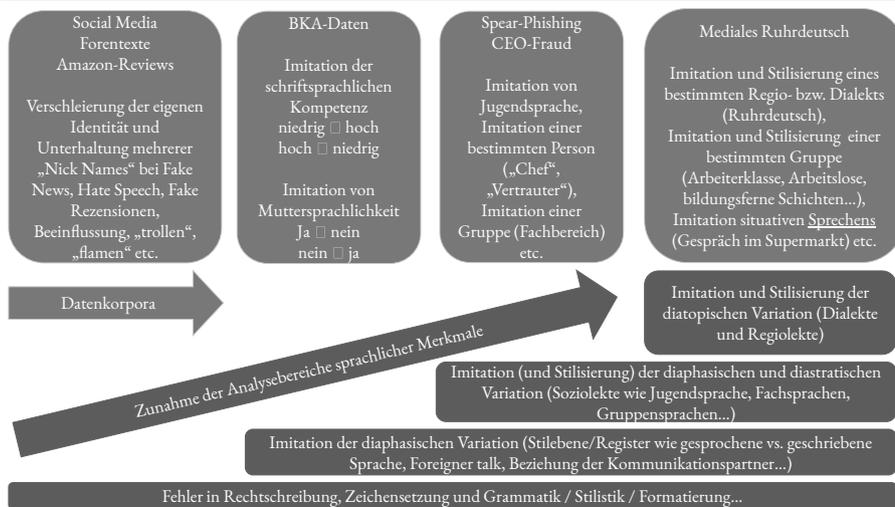


Abb. 3 | Analysebereiche sprachlicher Merkmale



tation einer geringeren oder höheren schriftsprachlichen Kompetenz. In den Bereichen Spear Phishing oder CEO-Fraud sind weitere Imitationen sprachlicher Varietäten möglich, wie z. B. die Imitation von Jugendsprache mit dem Gebrauch bestimmter Substandardkonstruktionen wie „yolo“ oder „gönn dir“, Grußformeln wie „tachkes“ bzw. „shalom“ (vgl. [11]) oder die Verwendung einer bestimmten Fachsprache.

**Stilisierung:** Die Stilisierung anderer Identitäten ist die komplexeste Verstellungsstrategie. Hier kann sich die/der Verfasser/ in eines Textes durchaus dessen bewusst sein, dass die Verstellung nicht der Stilistik einer bestimmten Gruppe oder Person entspricht. Die Verstellung ist stilisiert. Das bedeutet, dass sprachliche Merkmale sowohl qualitativ als auch quantitativ „übertrieben“ verwendet werden können. Ein Bereich, in dem diese Strategie eingesetzt wird, ist beispielsweise das medial stilisierte Ruhrdeutsch (vgl. [9]), bei dem sowohl Merkmale eines Regiolekts (zur gesprochenen Sprache im Ruhrgebiet vgl. [10]), bestimmter Gruppen (wie beispielsweise „Arbeitersprache“) mit Merkmalen situativen Sprachgebrauchs (z. B. „Gespräch im Supermarkt“) kombiniert eingesetzt und überstilisiert werden. Abb. 3 zeigt die Zunahme der Analysebereiche sprachlicher Merkmale in Abhängigkeit von verschiedenen Datenkorpora.

Obwohl die Formen schriftsprachlicher Verstellung und damit auch die Analysebereiche sprachlicher Merkmale bestimmten Datenkorpora wie in Abb. 3 zugeordnet werden können, ist der Gebrauch nicht ausschließlich auf diese Bereiche beschränkt. So bedient sich zwar das mediale Ruhrdeutsch grundsätzlich aller drei Formen, jedoch können beispielsweise auch in einem Erpresserbrief stilisierte Merkmale verwendet werden, um die eigene Identität möglichst effektiv zu verheimlichen. Aus diesem Grund ist die Betrachtung komplexerer Strategien auch für den Kernbereich der Forensischen Linguistik sehr vielversprechend.

### 3 Aspekte der Automatisierung

Die große Menge über das Internet veröffentlichter Daten verlangt automatisierte Verfahren, die die linguistische Analyse

unterstützen. Die technische Autorenerkennung stellt im Wesentlichen eine wissenschaftliche Disziplin dar, die untersucht, inwieweit Analysen von Textsammlungen hinsichtlich der Urheberschaft und möglicher biographischer Aussagen automatisiert werden können. Eine praktische Bedeutung erhält sie in der Kriminaltechnik bzw. IT-Forensik, um z. B. Erpresser- oder Bekenner schreiben und neuerdings auch Fake-News oder Hate-Speech zu untersuchen, wenn Meta-Daten keine Rückschlüsse auf die Urheberschaft zulassen. Eine andere Anwendung ist die Zuordnung anonymer historischer Textsammlungen, wenn man den Kandidatenpool eingrenzen kann. Eng verwandt mit der Autorenerkennung und von großer praktischer Bedeutung ist das Erkennen von Plagiaten.

Eng verwandt mit der Autorenerkennung und von großer praktischer Bedeutung ist das Erkennen von Plagiaten.

Mit Hilfe moderner Verfahren aus dem Bereich der künstlichen Intelligenz können Teilprozesse wie die Extraktion sprachlicher Merkmale automatisiert, aber auch Gesamtsysteme basierend auf Methoden des Deep Learning realisiert werden. Mit dem weltweiten Anstieg der Datenmenge sowie der wachsenden Rechenleistung erfreuen sich tiefe neuronale Netze immer größerer Beliebtheit. Sie werden in Verbindung mit großen Datensätzen genutzt, um Muster und Strukturen in diesen Daten in einem Lernvorgang abzuleiten. Mit jedem neuen Trainingsbeispiel bewertet das neuronale Netz sein Klassifikationsergebnis und stellt daraufhin seine Parameter neu ein [17]. Computergestützte Verfahren können somit eine Hilfestellung für die Akteure darstellen, um die linguistische Analyse auf Zweifelsfälle zu reduzieren.

Im Folgenden beschäftigen wir uns insbesondere mit der automatisierten Analyse der Autorschaft von Online-Texten und untersuchen, inwieweit eine Klassifikation mit hoher Zuverlässigkeit in Datenmaterial von inhärent großer Variabilität möglich ist. Aus technischer Sicht werden im Wesentlichen zwei Disziplinen unterschieden: die *Autorschafts-Attribution* (engl.: Authorship Attribution), sowie die *Autorschafts-Verifikation* (engl.: Authorship Verification).

#### 3.1 Attribution vs. Verifikation

Die *Autorschafts-Attribution* stellt eine klassische Aufgabe der Mustererkennung dar, mit dem Ziel, einem Dokument mit unbekannter Urheberschaft eine/n Kandidaten/in aus einer endlichen bekannten Menge von Personen zuzuordnen. Ein entscheidender Nachteil ist, dass alle in Frage kommenden Personen a priori bekannt sein müssen, d. h. wir müssen von jeder Person im Vorfeld genügend Textsammlungen zur Verfügung haben, um einen Klassifikator zu trainieren.

Die *Autorschafts-Verifikation* (oder *Textvergleich*, siehe Kapitel 2.1) wiederum untersucht zwei Texte bzw. Textsammlungen von zwei Personen hinsichtlich der Urheberschaft: Hier soll der trainierte Klassifikator entscheiden, ob zwei gegebene Texte von der gleichen Person verfasst wurden.

Im Folgenden behandeln wir die Autorschafts-Verifikation. Ein solches Modell besteht im Wesentlichen aus den folgenden Arbeitsschritten:

1. **Extraktion computerlinguistischer Merkmale:** Beide Texte werden in eine für den Computer sinnvolle Repräsentation überführt.
2. **Ähnlichkeitsbestimmung:** Auf Basis der extrahierten computerlinguistischen Merkmale entscheidet ein Klassifikator, ob bzw. inwieweit die Texte von der gleichen Person verfasst wurden. Der Klassifikator wird im Vorfeld anhand von Beispieltextsammlungen trainiert.

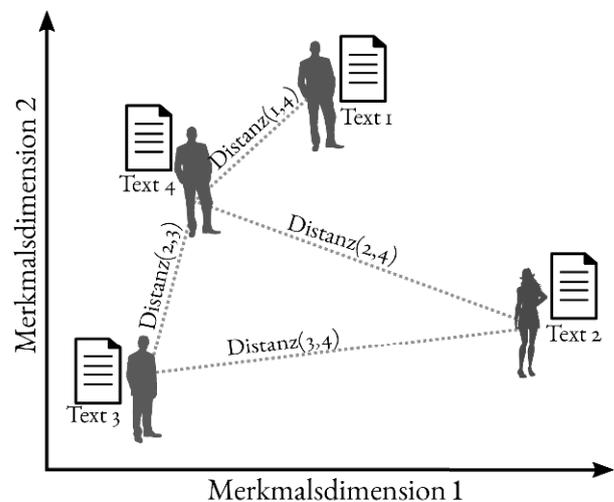
### 3.2 Extraktion computerlinguistischer Merkmale

Den entscheidenden Baustein eines technischen Systems stellt die automatische Extraktion von computerlinguistischen Merkmalen dar. In Abschnitt 2 wurde erläutert, dass die forensische Linguistik sprachliche Merkmale auf verschiedenen linguistischen Ebenen (z. B. Graphematik, Morphologie, Syntax, Orthographie etc.) untersucht. Im Gegensatz dazu ist eine äquivalente computerlinguistische Realisierung auf verschiedenen sprachlichen Ebenen weitaus komplexer und bis heute in dieser Form nicht realisiert. Computerlinguistische Merkmale werden in sogenannte Merkmalsvektoren strukturiert, wobei ein einzelner Eintrag in dem Vektor ein extrahiertes Merkmal darstellt. Die Abbildung 4 zeigt exemplarisch zwei computerlinguistische Merkmale von Texten dreier verschiedener Personen. In diesem Fall stellen die Merkmalsvektoren Punkte in einer Ebene dar. Für eine Ähnlichkeitsbestimmung ist es nun wünschenswert, dass die extrahierten Merkmale von Texten der gleichen Person (z. B. Text 1 und Text 4) ähnlich sind, d. h. eine geringe Distanz zueinander aufweisen. Die Merkmale von Texten von verschiedenen Personen (z. B. Text 2 und Text 4) sollen sich dagegen möglichst unähnlich sein, also eine größere Distanz zueinander haben. Es folgt nun die Ähnlichkeitsbestimmung. Hierbei kann u. a. ein Klassifikator eingesetzt werden, der während des Trainings die verschiedenen Abstände zusammen betrachtet, um dann zu ermitteln, bei welchem Abstand entschieden wird, ob die Texte von der gleichen Person verfasst wurden.

Wir unterscheiden zwei Arten von computerlinguistischen Merkmalen:

- **Stilometrische Merkmale:** Stilometrische Merkmale werden im Vorfeld definiert und dann automatisiert erhoben. Sie sind daher linguistisch interpretierbar. Ein einzelner Eintrag in diesem Merkmalsvektor beschreibt in der Regel quantitative Beobachtungen (i.d.R. relative Häufigkeiten) von z. B. Zeichen, Wörtern, Wort- und Satzlängen, Wortarten oder Phrasen in einem Text.
- **Neuronale Merkmale:** Im Gegensatz zu stilometrischen Merkmalen werden neuronale Merkmale nicht im Vorfeld definiert, sondern während des Trainings zusammen mit einem Klassifikator für die Ähnlichkeitsbestimmung gelernt. Die einzelnen Einträge im Merkmalsvektor werden auf Basis der Trainingsdaten durch das System eigenständig definiert und sind daher linguistisch nicht (direkt) interpretierbar. Allerdings werden wir in Kapitel 3.7 sehen, dass die Entscheidungsfindung unseres Systems auf Basis der neuronalen Merkmale graphisch darstellen werden kann.

Abb. 4 | Repräsentation von Texten durch Merkmalsvektoren



### 3.3 Neue Herausforderungen im digitalen Zeitalter

Im vorausgegangenen Abschnitt wurden zwei Arten zur Extraktion computerlinguistischer Merkmale eingeführt. Es stellt sich nun die Frage, welche Faktoren auf diesen Prozess einwirken, inwieweit dadurch eine Analyse der Autorschaft erschwert wird und welche Anforderungen daher an computerlinguistische Merkmale gestellt werden müssen. In Abb. 4 sind die Möglichkeiten der Beeinflussung graphisch dargestellt und werden im Folgenden genauer erläutert:

- **Schreibstil:** Als Schreibstil werden die in Abschnitt 2 erwähnten sprachlichen Merkmale wie z. B. Abweichungen von der Standardsprache definiert. Um Abweichungen von der Standardsprache quantitativ zu erfassen, müssen die zu untersuchenden Textsammlungen entsprechend lang genug sein. Der Schreibstil wird durch mögliche Verstellungsstrategien beeinflusst.
- **Genre:** Das Genre beschreibt die Textsorte. Mögliche Textsorten sind: Erpresserschreiben, Amazon-Rezensionen, Tweets, WhatsApp-Nachrichten, E-Mails an Kollegen/innen oder ein Anschreiben für eine Bewerbung. Je nach Genre verwenden wir ein anderes sprachliches Register, wodurch wir unterschiedliche computerlinguistische Merkmale erhalten. Für ein technisches System ist es nahezu unmöglich, eine Sammlung von WhatsApp-Nachrichten sinnvoll mit einem förmlichen Bewerbungsanschreiben zu vergleichen. Daher sollte ein Klassifikator jeweils nur auf ein Genre trainiert werden.
- **Thema:** Unter dem Begriff Thema verstehen wir den Inhalt bzw. die Botschaft, die eine Person den Lesern/innen mitzuteilen versucht. Das verwendete Vokabular ist abhängig von dem jeweiligen Thema und kann stark variieren. Zudem wird das Vokabular durch die verfasste Meinung beeinflusst. Soll eine positive oder negative Meinung zum Ausdruck gebracht werden? Ein themenspezifisches Vokabular übt einen erheblichen Einfluss auf quantitative Merkmale aus.

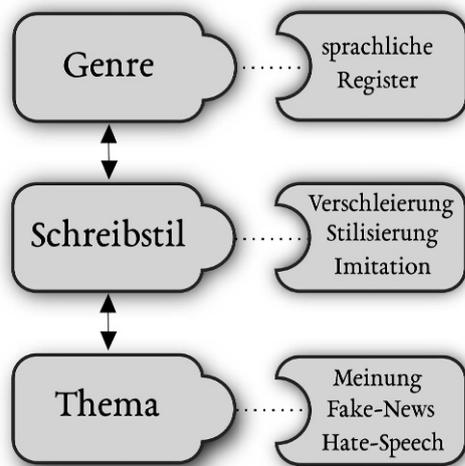
Die automatisierte Extraktion computerlinguistischer Merkmale wird bei Texten aus sozialen Medien durch zusätzliche Faktoren deutlich erschwert: *Digitale* Texte sind in der Regel verhältnismäßig kurz (deutlich weniger als 1000 Wörter). Gleichzeitig lassen sich eine hohe Variabilität bzgl. des Kontextes und ein si-

gnifikanter Unterschied zwischen dem Standard- und dem Online-Sprachgebrauch feststellen, weshalb vorangestellte Verarbeitungsschritte, wie z. B. die Detektion von Satzgrenzen oder die Erkennung von Eigennamen und Wortarten fehlerhafte Ergebnisse produzieren, die durch eine Fehlerfortpflanzung in der Verarbeitungskette dann auch Auswirkungen auf die Merkmalsextraktion und letztlich auf die Erkennungsgenauigkeit haben. Aus diesem Grund haben wir drei grundlegende Anforderungen formuliert, die computerlinguistische Merkmale im Umgang mit digitalen Texten im Idealfall erfüllen sollen:

- **Unterscheidbarkeit:** Computerlinguistische Merkmale sollen unterscheidungsrelevante Informationen repräsentieren, die eine Trennung bzgl. der Autorschaft erlauben. Irrelevante Informationen (u. a. Zitate, URLs) sollen statistisch nicht erfasst werden. Insbesondere sollen mit Hilfe von computerlinguistischen Merkmalen verschiedene Schreibstile unterschieden werden, es sollen nicht die spezifischen Charakteristika einzelner Autoren auswendig gelernt werden.
- **Invarianz:** Computerlinguistische Merkmale sollen unabhängig vom Thema, der zum Ausdruck gebrachten Meinung oder dem verwendeten sprachlichen Register sein. Eine Entscheidungsfindung soll auf Basis des Schreibstils und nicht auf Basis von inhaltlichen Aussagen erfolgen.
- **Robustheit:** Computerlinguistische Merkmale sollen unterscheidbar bleiben, auch wenn Strategien eingesetzt werden, um die eigene Autorschaft zu verschleiern.

Wir werden später auf die Unterscheidbarkeit und Invarianz noch genauer eingehen und quantitativ sowie qualitativ evaluieren, inwieweit unser System kontextsensitiv ist.

Abb. 5 | Beeinflussung der Extraktion computerlinguistischer Merkmale



### 3.4 Das Verfahren AdHominem

In diesem Abschnitt präsentieren wir unser Verfahren, das mit den zuvor beschriebenen Herausforderungen umzugehen vermag. Die Aufgaben des Systems sind sowohl die Extraktion von linguistischen Merkmalen als auch die Ähnlichkeitsbestimmung, d. h. zu entscheiden, inwieweit sich die Schreibstile für zwei verschiedene Amazon-Rezensionen ähneln.

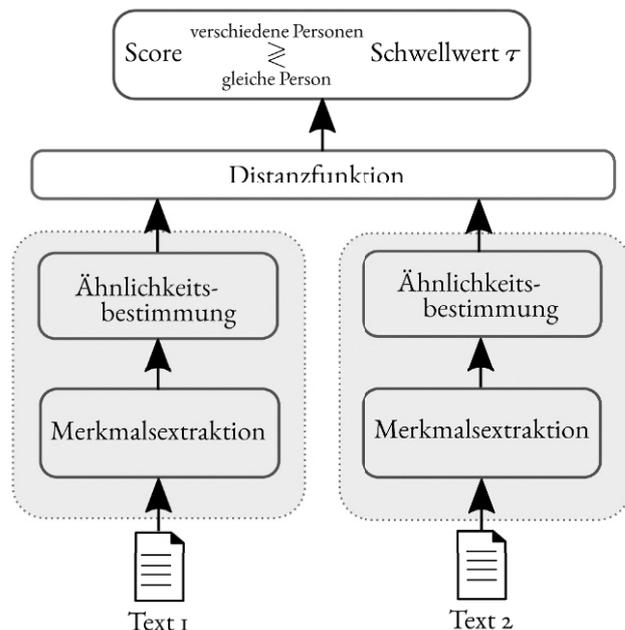
Unser Verfahren haben wir AdHominem getauft. Das Akronym steht für *Attention-based Deep Hierarchical Convolutional siamese bidirectional recurrent neural-network Model*. AdHominem ist schematisch in Abb. 6 dargestellt. Eine detaillierte Beschreibung des Verfahrens findet sich in [15].

Stilometrische Merkmale liefern erst bei ausreichend langen Texten zuverlässige Ergebnisse, d. h. wir benötigen i.d.R. eine Textsammlung, in der einzelne Texte entsprechend lang sein müssen. Wie sollen wir also umgehen mit relativ kurzen Texten, die weniger als 1000 Wörter enthalten? AdHominem wurde deswegen im Vergleich zu [16] entwickelt, um mit einer ausreichend großen Menge von Textpaaren bestehend aus relativ kurzen Texten arbeiten zu können.

Wie in Abb. 6 dargestellt, enthält AdHominem im Kern eine siamesische Netzwerktopologie, d. h. wir haben zwei separate neuronale Netze, welche die gleiche hierarchische Struktur und die gleichen gelernten Parameter aufweisen. Aus Sicht der Implementierung stellt also das linke neuronale Netz eine exakte Kopie des rechten Netzes dar. In der Literatur hat sich für diese Form der Begriff *siamesisches Netzwerk* etabliert.

AdHominem ist in verschiedene Module unterteilt: Im ersten Schritt werden computerlinguistische Merkmale aus den beiden Eingabetexten extrahiert (Modul Merkmalsextraktion). Jedes Dokument stellt eine hierarchische Struktur bestehend aus Sätzen, Wörtern und Buchstaben dar. In einem dreistufigen Verfahren werden daher Merkmale auf Buchstaben-, Wort- und Satzebene extrahiert. Diese werden während des Trainings anhand von vielen Beispieltexten erlernt. Im zweiten Schritt folgt die Ähnlichkeitsbestimmung auf Basis dieser Merkmale. Dieses Modul wird ebenfalls anhand von Beispieltexten trainiert. Eine Distanzfunktion ermittelt schließlich eine Zahl (Score). Liegt dieser ermittelte Score unter einem Schwellenwert, entscheidet das System, dass beide Texte von der gleichen Person verfasst wurden. Andernfalls entscheidet sich das System dafür, dass zwei verschiedene Personen beteiligt waren. Der Schwellenwert wird vor dem Training festgelegt und danach nicht mehr verändert.

Abb. 6 | AdHominem



### 3.5 Korpuserstellung auf Basis einer Big-Data-Textsammlung

Für eine quantitative Analyse von Klassifikatoren hinsichtlich der Beeinflussung computerlinguistischer Merkmale benötigen wir ein großes annotiertes Korpus. Das in unserem Experiment genutzte Korpus basiert auf einer frei verfügbaren Textsammlung [19] von Millionen von Amazon-Rezensionen. Aus dieser Datenmenge haben wir eine für unser Anliegen entsprechende Teilmenge wie folgt gebildet:

- Rezensionen mit weniger als 80 und mehr als 1000 Wörtern werden nicht berücksichtigt. Als Ergebnis haben wir 9.052.606 Bewertungen, die von 784.649 Personen im Zeitraum von 1996 bis 2014 verfasst wurden. Eine Rezension besteht im Durchschnitt aus 283 Wörtern.
- Es gibt insgesamt 24 verschiedene Kategorien für Themen, z. B. Bücher, Garten, Elektronik etc.
- Wir bilden zufällig Paare von Rezensionen, die hinsichtlich der Autorschaft sowie der Kategorie annotiert sind. Jedes Paar enthält ein Tupel der Form  $l=(a,c)$ . Der Wert von  $a$  gibt an, ob beide Rezensionen von derselben Person verfasst wurden ( $a=1$ ) oder nicht ( $a=0$ ). Der Wert von  $c$  gibt an, ob die beiden Rezensionen dasselbe Thema ( $c=1$ ) oder verschiedene Themen ( $c=0$ ) behandeln.
- Je mehr Trainingsbeispiele das neuronale Netz erhält, desto besser kann es generalisieren, d. h. es verbessert seine Fähigkeit, auf zuvor noch nicht gesehene Daten zu reagieren. Daher werden sämtliche Trainingsbeispiele, in unserem Fall Paare von Amazon-Rezensionen, in vielen Runden (Epochs) in geänderter Reihenfolge mehrmals durchlaufen. Um die Heterogenität der Daten zu erhöhen, kombinieren wir zusätzlich die Paare nach jeder Iteration neu.

Hierbei machen wir die folgenden Annahmen:

- Ein Account wird von genau einer Person benutzt. D. h. eine Rezension wurde von genau einer Person verfasst.
- Eine Person besitzt maximal einen Account. D. h. hinter zwei User-ID s verbirgt sich nie dieselbe Person.
- Es sind keine Fake-Accounts bzw. Fake-Rezensionen vorhanden.
- Es sind keine Verstaltungen, d. h. Stilisierungen, Imitationen oder Verschleierungen zu erwarten.

Wir können diese Annahmen bei einem Korpus dieser Größenordnung nicht garantieren. Allerdings ist anzunehmen, dass Abweichungen in verschwindend geringer Menge auftreten und somit von unserem System als nicht signifikant erachtet werden.

### 3.6 Evaluierung

In Tabelle 1 sind die durchschnittlichen Fehlerraten inklusive Standardabweichung aufgeführt. Für unseren Vergleich wählten wir zwei Algorithmen (AVEER und GLAD [18]), die stilometrische Merkmale, aber unterschiedliche Verfahren für die Ähnlichkeitsbestimmung verwenden. In der ersten Zeile sehen wir die Ergebnisse gemittelt für alle Annotationen, d. h. für ( $a=1, c=1$ ), ( $a=1, c=0$ ), ( $a=0, c=1$ ) und ( $a=0, c=0$ ).

Wie in der ersten Zeile der Tabelle zu sehen ist es mit AdHominem möglich, die Fehlerrate etwa zu halbieren. Die Zeilen 2 bis 5 zeigen die Ergebnisse in Abhängigkeit der Annotationen. Hier zeigen alle drei Verfahren eine starke Abhängigkeit hinsichtlich des Kontextes. In den Zeilen 2 und 3 sehen wir die Fehlerraten für

Paare von Rezensionen, die von der selben Person verfasst wurden ( $a=1$ ). Im Vergleich der beiden Zeilen ist zu erkennen, dass sich die Fehlerrate um 5 bis 10% erhöht, sobald ein/e Autor/in in den beiden Texten über verschiedene Themen schreibt. Im Gegensatz dazu sehen wir in Zeilen 4 und 5 die Ergebnisse für Rezensionen, die von verschiedenen Personen verfasst wurden ( $a=0$ ). Vergleichen wir Zeilen 4 und 5 miteinander, ist zu erkennen, dass sich die Fehlerraten nun um mehr als 10% erhöhen, wenn die verschiedenen Personen in ihren Rezensionen das gleiche Thema behandeln.

**Tab. 1 | Vergleich der durchschnittlichen Fehlerrate (mit Standardabweichung) verschiedener Algorithmen**

Annotation	Fehlerrate in %		
	AdHominem	AVEER	GLAD
$\forall l = (a, c)$	$14.65 \pm 0.11$	$30.13 \pm 0.13$	$27.13 \pm 0.01$
$a = 1, c = 1$	$10.48 \pm 1.12$	$25.63 \pm 0.25$	$24.27 \pm 0.23$
$a = 1, c = 0$	$15.40 \pm 0.89$	$35.88 \pm 0.53$	$35.97 \pm 0.57$
$a = 0, c = 1$	$22.65 \pm 1.58$	$36.37 \pm 0.54$	$30.28 \pm 0.54$
$a = 0, c = 0$	$10.09 \pm 0.68$	$22.64 \pm 0.37$	$18.02 \pm 0.41$

### 3.7 Visualisierung der Entscheidungsfindung

Da die englischsprachigen Rezensionen bereits durch das automatische System AdHominem untersucht wurden, sollen nun einige Beispiele für die linguistischen Merkmale aus diesem Korpus gegeben werden. Ein Vorteil von AdHominem besteht darin, dass das System bestimmte Textteile, die signifikant für seine Entscheidung sind, ob zwei Texte von einer/einem gleichen Autor/in verfasst wurden oder nicht, entsprechend visualisiert. Das System bietet somit Anhaltspunkte für quantitative und qualitative Einschätzungen, die im Bereich der Forensischen Linguistik neue Forschungsansätze bieten.

So werden bestimmte Textteile stufenlos in Rottönen nach Signifikanz für das System markiert. Je intensiver die Färbung, desto interessanter ist der eingefärbte Bereich für das System.

Als quantitative Analyse bietet sich an, wie häufig bestimmte Textteile vom System als „signifikant“ markiert werden. Es sind sowohl prozentuale und als auch absolute Angaben zur Frequenz bestimmter Merkmale sinnvoll. Ein Beispiel wäre, welche Satzzeichen wie häufig und auch im Verhältnis zu anderen Satzzeichen markiert werden, d. h. also für das System signifikant für die Entscheidung über die Autorschaft sind.

Weiter kann beobachtet werden, dass es bei der Entscheidung des Systems, ob es sich um dieselbe/denselben Autorin/Autor handelt, eine Rolle spielt, an welcher Position eines Textes bestimmte Textteile markiert werden.

Neben dem Scoring-Wert erhalten wir als Ausgabe für jedes Wort und für jeden Satz sogenannte *Attention-Gewichte*. Diese beschreiben den Grad der Relevanz der Wörter oder Satzzeichen für die neuronalen Merkmale. Im Folgenden präsentieren wir eine kleine Auswahl an Markierungen, die bestimmten Kategorien und Unterkategorien zugeordnet wurden. Hierbei stellen die roten Markierungen die Attention-Gewichte dar. Je stärker die Intensität, desto höher ist die Relevanz:

**Zeichensetzung:**

- 1 Satzzeichenhäufung:

Beispiel: ...  
I just read the chapter on Generosity ... and it was PHENOMENAL !

- 2 Sonderzeichen:

Beispiel: &  
[...] if you need to tweak & bend them

- 3 Fehlendes Leerzeichen:

Beispiel: book.s vs. books.s  
I highly recommend this book.s Stars

**Buchstaben:**

- 4 Buchstabenverwechslung:

Beispiel: deffinatly vs. definitely  
Ice Cube deffinatly has a style that is unimatatable .

- 5 Fehlender Buchstabe:

Beispiel: clasics vs. classics  
[...] he went on to put out some hip hop clasics .

- 6 Überzähliger Buchstabe / Hinzugefügter Buchstabe:

Beispiel: amazone vs. Amazon/amazon  
i searched a lot of this kind of high tech light around amazone

**Groß- und Kleinschreibung:**

- 7 Klein- statt Großschreibung:

Beispiel: i vs. I  
[...] i m glad i picked this .

- 8 Groß- statt Kleinschreibung:

Beispiel: Stars vs. stars  
I highly recommend this book.s Stars

- 9 Durchgehende Großschreibung:

Beispiel: TOTALLY AWESOME  
Uncaged is TOTALLY AWESOME !!!!!

**Getrennt- und Zusammenschreibung:**

- 10 Fehlerhafte Zusammenschreibung:

Beispiel: ripoff vs. rip off/rip-off  
[...] universe of camelot that is a disney ripoff [...]

- 11 Fehlerhafte Getrenntschreibung:

Beispiel: story teller vs. storyteller  
JJ Knight is an awesome story teller [...]

**Abkürzungen und Akronyme:**

- 12 Akronym:

Beispiel: OMG  
OMG , someone finally figured it out !

- 13 Ungewöhnliche Abkürzung:

Beispiel: def vs. definitely  
[...] and that track is def tight

**Diatopische Variationen & Fremdsprachen**

- 14 Britisches vs. amerikanisches Englisch :

Beispiel: favourite (BE) vs. favorite (AE)  
[...] this version is undoubtedly my favourite .

**Stilistik:**

- 15 Umgangssprachlicher Ausdruck:

Beispiel: thingamajig  
[...] percentage at the bottom thingamajig

- 16 Alternative Schreibung:

Beispiel: frikkin vs. freaking  
[...] , its a frikkin tv show .

- 17 Neologismus:

Beispiel: cartoonish  
But , Lily feels cartoonish to me , [...]

**Syntax:**

- 18 Verb an 1. Stelle im Aussagesatz (fehlendes Pronomen bzw. Substantiv)

Beispiel: [I] Will  
Will definitely read more of her books .

## 4 Zusammenfassung und Ausblick

In diesem Beitrag präsentierten wir den Zusammenhang verschiedener Verstellungsstrategien und unser Verfahren AdHominem, das die Extraktion computerlinguistischer Merkmale sowie die Ähnlichkeitsbestimmung in einem System integriert. Die gewählte Topologie ist geeignet für die Verarbeitung von kurzen Texten. Im Gegensatz zu klassischen Methoden auf Basis von stilometrischen Merkmalen benötigt unsere Deep-Learning-Methode hierfür große Textsammlungen für das Training.

Die Evaluierung in Abschnitt 3.6 zeigt sehr deutlich, dass Verfahren basierend sowohl auf stilometrischen als auch auf neuronalen Merkmalen eine hohe Kontextsensitivität aufweisen. Insgesamt können wir aber hier mit unserer Deep-Learning-Architektur die Fehlerrate halbieren. Zudem sind eine Visualisierung und eine (vorsichtige) Interpretation der Ergebnisse anhand der Attention-Gewichte möglich.

Zukünftige Schritte umfassen insbesondere zwei Ansätze hinsichtlich der beschriebenen Störfaktoren, welche die computerlinguistischen Merkmale beeinflussen: Zum einen wollen wir durch ein geeignetes Training die Kontextsensitivität verringern. Zum anderen werden die Architektur erweitert und das Training angepasst, um AdHominem robust gegenüber Verstellungsstrategien zu machen.

Die linguistische Analyse der Merkmale verschiedener Korpora liefert Erkenntnisse darüber, wie verschiedene Dimensionen schriftsprachlicher Verstellung funktionieren, voneinander abzugrenzen sind und wo Gemeinsamkeiten und Unterschiede verschiedener Textsorten und Medien liegen. Es konnten unterschiedliche linguistische Analysebereiche auf die erstellten Korpora angewendet werden. Die Ergebnisse können sowohl für Textanalysen als auch für Textvergleiche herangezogen werden.

Die Analysen von Merkmalen und deren Einordnung liefern zusammen mit der Interpretation der visuellen Repräsentationen aufschlussreiche Erkenntnisse bei der Weiterentwicklung von AdHominem. Die Analyse der Attentions bietet wiederum Anhaltspunkte für neue Formen qualitativer und quantitativer Analysen im Bereich der forensischen Linguistik. Die stufenlose Markierung bestimmter Textteile lässt Aussagen darüber zu, welche Merkmale mit welcher Signifikanz an welcher Position eines Textes für ein automatisches System bedeutsam sind, wenn es dar-

um geht, Aussagen darüber zu treffen, ob zwei (oder mehr) Texte von einer/einem Autorin/Autoren verfasst wurden oder nicht.

## Literatur

- [1] S. Bredthauer (2014): Verstellungen in inkriminierten Schreiben: Eine linguistische Analyse verstellten Sprachverhaltens in Erpresserschreiben und anderen inkriminierten Texten. Köln: Kölner Wissenschaftsverlag.
- [2] M. Coulthard (2010): A Handbook of Forensic Linguistics. London: Routledge.
- [3] C. Dern (2009): Autorenerkennung. Stuttgart, München etc.: Boorberg.
- [4] R. H. Drommel (2016): Sprachprofilung – Grundlagen und Fallanalysen zur Forensischen Linguistik. Berlin: Frank & Timme.
- [5] S. Ehrhardt (2018): Authorship attribution analysis. In: Visconti, Jacqueline (ed.): Handbook of Communication in the Legal Sphere. Berlin/Boston: de Gruyter, 169–200.
- [6] E. Fobbe (2014): Fingierte Lernersprachen. Strategien der muttersprachlichen Fehlerproduktion im Dienste der Verstellung. In: Zeitschrift für germanistische Linguistik. Deutsche Sprache in gegenwart und Geschichte. Heft 42.2. Berlin: De Gruyter. S.196–222.
- [7] E. Fobbe (2011): Forensische Linguistik. Tübingen: Narr.
- [8] E. Fobbe (2006): *Foreigner talk* als Strategie. Zur Fehlergenese in Erpresserbriefen. In: Busch, Albert/Sigurd Wichter (Hgg.): Wissenstransfer: Erfolgskontrolle und Rückmeldungen aus der Praxis, Frankfurt/M., Bern u.a.: Lang. S. 149–165.
- [9] S. Hessler (2018): Echt ruhrdeutsch? Kult, Klischee und Sprache in Filmen des östlichen Ruhrgebiets. In: Begleitband der Ausstellung „Revierfolklore – Zwischen Heimatstolz und Kommerz“ des LWL-Industriemuseums Zeche Hannover. Essen: Klartext. S. 129–141.
- [10] S. Hessler, D. Pottmann (2017): Merkmale der gesprochenen Sprache des mittleren und östlichen Ruhrgebiets. In: Sprache & Sprachen. (=Zeitschrift der Gesellschaft für Sprache und Sprachen. Band 47.) S. 1–18.
- [11] S. Hessler (2016): Temporalität und die Verwendung von definierten Grußformeln wie hallöchen und tschüsschen am Beispiel der gesprochenen Sprache des Ruhrgebiets. In: ZDL. Stuttgart: Steiner. (= Zeitschrift für Dialektologie und Linguistik. Band 83.) S. 1–26.
- [12] S. Hessler (2012): Temporalität in ausgewählten Substandardkonstruktionen. Zur Verwendung von am-Progressiv-Formen und diminuierten Grußformeln in der gesprochenen Sprache des Ruhrgebiets. Master-Arbeit an der Ruhr-Universität Bochum.
- [13] McMenamin, Gerald (2003): Forensic Linguistics: Advances in Forensic Stylistics. Boca Raton: CRC Press.
- [14] Shuy, Roger W. (2016): The Language of Fraud Cases. Oxford: Oxford University Press.
- [15] B. Boenninghoff, R. M. Nickel, S. Zeiler and D. Kolossa (2019): Similarity Learning for Authorship Verification in Social Media, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom.
- [16] A. Rocha et al. (2017): Authorship Attribution for Social Media Forensics, IEEE Transactions on Information Forensics and Security, vol. 12, no. 1.
- [17] P. Shrestha, P. Sierra, F. González, M. Montes, P. Rosso, T. Solorio (2017): Convolutional Neural Networks for Authorship Attribution of Short Texts, Conference of the European Chapter of the Association for Computational Linguistics (EACL), Valencia, Spain.
- [18] O. Halvani, C. Winter and L. Graner (2019): Assessing the Applicability of Authorship Verification Methods, International Conference on Availability, Reliability and Security (ARES), Canterbury, United Kingdom.
- [19] R. He, J. McAuley (2016): Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering, WWW.

# Neuerscheinung



A. Sieber  
**Dialogroboter**

Wie Bots und künstliche Intelligenz Medien und Massenkommunikation verändern

2019, VII, 228 S. 31 Abb., 21 Abb. in Farbe. Brosch.

€ (D) 22,99 | € (A) 23,63 | \*sFr 25,50

ISBN 978-3-658-24392-0

€ 16,99 | \*sFr 20,00

ISBN 978-3-658-24393-7 (eBook)

- Erste systematische Darstellung des Phänomens Bots und Sprachdialogsysteme
- Wie Sprachdialogsysteme funktionieren und wie sie gebaut werden
- Bietet wissenschaftliche Grundlagen für die allgemeine Diskussion

Technologien wie künstliche Intelligenz und Natural Language Programming werden zu Auslösern der sogenannten „Dialogwende“. Darunter versteht dieses Buch die massenweise Verbreitung von autonom sprechenden Sprachdialogsystemen und automatischen Sprachassistenten.

## Ihre Vorteile in unserem Online Shop:

Über 280.000 Titel aus allen Fachgebieten | eBooks sind auf allen Endgeräten nutzbar |  
Kostenloser Versand für Printbücher weltweit

€ (D) sind gebundene Ladenpreise in Deutschland und enthalten 7 % für Printprodukte bzw. 19 % MwSt. für elektronische Produkte. € (A) sind gebundene Ladenpreise in Österreich und enthalten 10 % für Printprodukte bzw. 20 % MwSt. für elektronische Produkte.  
Die mit \* gekennzeichneten Preise sind unverbindliche Preisempfehlungen und enthalten die landesübliche MwSt. Preisänderungen und Irrtümer vorbehalten.

Jetzt bestellen auf [springer.com/informatik](http://springer.com/informatik) oder in der Buchhandlung

Part of **SPRINGER NATURE**