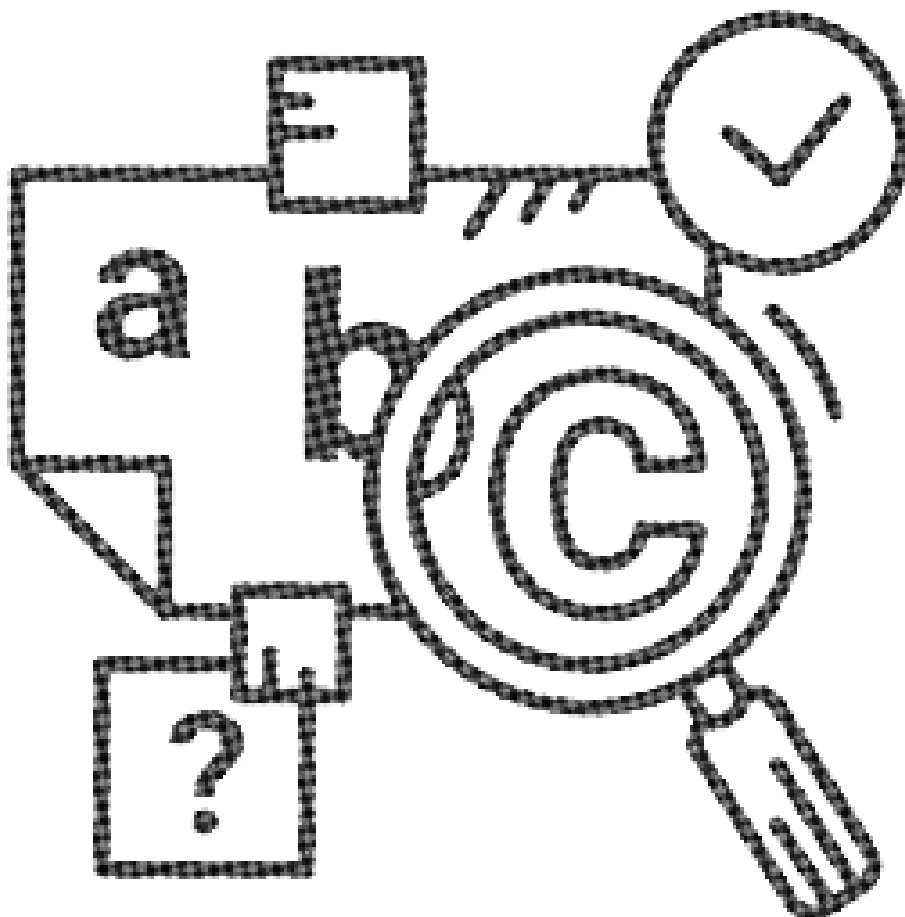


Wer schreibt denn da?

von Zündlumpen #76

2020



Ein kleiner Überblick über Methoden der modernen Forensischen Linguistik zur Autorschaftsbestimmung.

Der folgende Artikel versucht aus einer nicht fachlichen Perspektive einen Überblick zu geben und eine entsprechende Einordnung vorzunehmen. Es gibt einige wissenschaftliche Publikationen zu diesem Thema, die für eine bessere Einschätzung ausgewertet werden könnten. Es geht mir hier aber vor allem darum, das Thema einmal aufzuwerfen und nicht darum, eine fundierte und abschließende Betrachtung zu liefern. Wenn du also irgendetwas besser weißt, dann immer her mit den Informationen!

Spuren vermeiden, die einer später einmal – vielleicht noch nach Jahren und Jahrzehnten – zum Verhängnis werden könnten, das dürfte wohl für die Meisten von Interesse sein, die ab und an zur Tat schreiten und dabei in Konflikt mit dem Gesetz geraten. Fingerabdrücke vermeiden, DNA-Hinterlassenschaften vermeiden, Schuhabdrücke und Textilfaser-Spuren vermeiden oder zumindest getragene Kleidung im Anschluss entsorgen, Videoaufnahmen vermeiden, Werkzeugspuren vermeiden, Aufzeichnungen jeder Art vermeiden, Observationen erkennen usw., all das dürfte dabei zumindest mehr oder weniger jeder, die des öfteren Verbrechen begeht und sich dabei vor Identifizierung schützen will, ein Anliegen sein. Aber wie steht es mit jenen Spuren, die oft erst im Nachhinein eines Verbrechens aus dem Drang heraus, die eigene Tat wenigstens anonym oder auch unter Verwendung eines wiederkehrenden Pseudonyms zu erklären, entstehen? Beim Verfassen und Publizieren eines Communiqués oder eines Bekenner*innenschreibens?

Mein Eindruck ist, dass diesen Spuren trotz einer rasanten technologischen Entwicklung der Analysekapazitäten in vielen Fällen keine besondere Aufmerksamkeit gewidmet wird. Das kann Absicht sein, Nachlässigkeit oder auch ein Kompromiss aus miteinander konkurrierenden Bedürfnissen. Ohne hier einen allgemeinen Vorschlag zum Umgang mit diesen Spuren unterbreiten zu wollen – das muss schließlich jede für sich wissen –, möchte ich vor allem skizzieren, mit welchen Methoden die Ermittlungsbehörden in Deutschland und anderswo derzeit (wahrscheinlich) arbeiten, was grundsätzlich möglich scheint und was in Zukunft möglich werden könnte.

Vielleicht sollte ich vorab noch bemerken, dass freilich alles oder zumindest das allermeiste, was ich hier vorstelle wissenschaftlich ebenso wie juristisch umstritten ist. Ich bin auch weniger an der juristischen Verwertbarkeit von Sprachanalysen interessiert – und an der wissenschaftlichen sowieso nicht –, als daran, ob es plausibel erscheint, dass diese Ermittler*innen einer auf die Spur bringen, denn selbst wenn eine Spur gerichtlich nicht verwertbar ist, so kann es dennoch dazu führen, dass diese zur Ermittlung einer anderen, verwertbaren Spur führt.

Autorenerkennung beim BKA

Das Bundeskriminalamt unterhält eigenen Angaben zufolge eine Abteilung, die sich der Ermittlung der Autor*innenschaft bei Texten widmet. Im Fokus stehen dabei Texte mit einem Bezug zu Straftaten wie Bekenner*innenschreiben, aber auch „Positionspapiere“ unter anderem aus dem „linksextremistischen Spektrum“. Alle gesammelten Texte werden aufbereitet durch sprachwissenschaftliche Untersuchungen in einer sogenannten Tatschreibensammlung erfasst und sind mit dem Kriminaltechnischen Informationssystem Texte (KISTE) vergleich- und durchsuchbar. Den Angaben des BKA zufolge werden die Texte unter anderem klassifiziert nach den folgenden biografischen Merkmalen ihrer (vermeintlichen) Autor*innen: Herkunft, Alter, Bildung und Tätigkeit.

Alle eingehenden Texte werden zudem mit bereits erfassten Texten verglichen, um zu bestimmen, ob mehrere Texte möglicherweise von der gleichen Autor*in verfasst wurden.

Im Rahmen fallspezifischer Ermittlungen können die gespeicherten Texte zudem mit Texten, deren Autor*innenschaft bekannt ist verglichen werden, um zu bestimmen, ob diese von der gleichen Autor*in verfasst wurden, oder ob dies ausgeschlossen werden kann.

Soweit die offiziellen Angaben des BKA zu dieser Abteilung. Was bedeutet das in der Praxis?

Ich denke, dass man davon ausgehen kann, dass zumindest alle Bekenner*innenschreiben in dieser Datenbank erfasst werden und daraufhin analysiert werden, ob von der/den gleichen Autor*in(en) noch weitere Bekenner*innenschreiben vorhanden sind. Aber die Feststellung, dass auch „Positionspapiere“ erfasst werden, lässt noch weitere Schlüsse zu: Zumindest erscheint es möglich, dass neben Texten

mit strafrechtlicher Relevanz auch andere Texte eingespeichert werden, die einer bestimmten Szene zugeordnet werden. Beispielsweise Texte aus entsprechenden Zeitungen, Erklärungen von politischen Gruppen/Organisationen, Aufrufe, Blogbeiträge, usw. Im schlimmsten Fall würde ich also davon ausgehen, dass alle publizierten Texte auf bekannten „linksextremistischen“ Webseiten (da ist es schließlich recht einfach, an diese ranzukommen), sowie den Ermittlungsbehörden interessant erscheinende Texte aus Printpublikationen in diese Datenbank eingespeist werden.

Das würde bedeuten, dass dem BKA zu jedem Bekenner*innenschreiben ein Cluster aus Texten mit vermeintlich gleicher Autor*innenschaft vorliegt. Diese können dabei aus anderen Bekenner*innenschreiben bestehen sowie eben auch aus jenen Texten, die sonst noch so in die Datenbank eingespeist wurden. Neben Tatserien können so also auch weitere Hinweise auf Täter*innen gewonnen werden, etwa Pseudonyme, Gruppenbezeichnungen – oder schlimmstenfalls Namen – unter denen eine Verfasser*in eines Bekenntnisses andere womöglich andere Texte verfasst hat, aber je nach Text auch alle möglichen anderen Informationen, die dieser liefert, darunter häufig Hinweise auf Wohn- und Wirkungsort einer Person, thematische Schwerpunkte, biografische Charakteristika, Bildungsweg, usw. Allesamt Informationen, die mindestens dazu genutzt werden können, um den Kreis der Verdächtigen einzuschließen.

Was bei all dem noch unklar bleibt ist, welche weiteren Vergleichsproben das BKA möglicherweise vorhält. Von den meisten Personen gibt es sicher eine ganze Reihe Texte, auf die Ermittlungsbehörden Zugriff haben (könnten) und die im Falle eines Verdachts oder möglicherweise zum Teil auch vorsorglich – wenn eine Person etwa mit einem Eintrag wie „Gewalttäter linksextrem“, etc. bekannt ist – in die Datenbank eingespeist werden könnten. Das kann alles sein, wo dein Name drunter steht, vom Schreiben an eine Behörde bis hin zu einem Leserbrief in der Zeitung unter deinem Namen. Ich will hier absichtlich nur die offensichtlichsten Quellen nennen, um nicht versehentlich den Ermittlungsbehörden die entscheidende Inspiration zu verschaffen, aber ich bin sicher du kannst für dich selbst beantworten, welche Texte von dir zugänglich sein könnten. Gelingt es den Profiler*innen des BKA erst einmal den Verdächtigenkreis auf ein spezifisches Charakteristikum einzugrenzen, das den Abgleich mit massenhaft vorhandenen Textproben ermöglicht (Wenn beispielsweise davon ausgegangen wird, dass ein*e Wissenschaftler*in einer bestimmten Disziplin für ein Schreiben verantwortlich ist, könnten alle Publikationen in diesem Fachbereich als Vergleichsproben herhalten. Das wäre zum Beispiel eine mögliche (Teil-)Erklärung dafür, wie das mit Andrej Holm im Verfahren gegen die militante Gruppe gelaufen sein könnte, zumindest wenn man unterstellt, dass das BKA nicht nur nach „Gentrifizierung“ geogoglet hat), so halte ich es durchaus für möglich, dass solche Analysen auch durchgeführt werden.

Methoden der Autorenerkennung und des Autoren-Profilings

All das betrachtet aber nur, was das BKA von sich behauptet zu können und führt diese Überlegungen weiter. Aber wie funktioniert denn nun eigentlich die Autorenerkennung, bzw. das Autorenprofilung?

Wer kennt sie nicht, die Angst davor, dass eine*n vielleicht der*die Deutschlehrer*in enttarnen wird, nachdem auf der Toilette eine Spottdichtung über eine*n Lehrer*in aufgetaucht ist und sich die ganze Schule darüber lustig macht, wie man nur „Leerer“ statt „Lehrer“ schreiben könne. Aber glücklicherweise ist dann doch das gesamte Deutschkollegium darauf hereingefallen, das Narrativ vom Fehler zu übernehmen und die Augen vor einem nur allzu treffenden Wortspiel zu verschließen. Die Forensische Linguistik scheint doch ein wenig Übung oder zumindest eine kriminalistische Motivation zu erfordern, wer weiß. Jedenfalls war die Fehleranalyse, von der wohl die meisten schon einmal gehört haben dürften, zusammen mit der Stilanalyse einem Werbeartikel der Sprachbullin Christa Baldauf zufolge um 2002 herum eines der wichtigsten Analyseinstrumente des BKA. Rechtschreibfehler, Grammatikfehler, Interpunktion, aber auch Tippfehler, Neue oder Alte Rechtschreibung, Hinweise auf Tastatureigenheiten, usw., all das dient den Sprachbullen dazu, Hinweise auf den*die Autor*in zu sammeln. Wenn ich beispielsweise „muß“ statt „muss“ schreibe, könnte das ein Hinweis darauf sein, dass ich zu Schulzeiten einige der jüngeren Rechtschreibreformen nicht mehr mitbekommen habe. Wenn ich dagegen Begriffe, die man der Rechtschreibung zufolge mit „ß“ schreibt, ständig mit „ss“ schreibe, könnte das bedeuten, dass auf

meiner Tastatur kein „ß“ vorhanden ist. Wenn ich zum Beispiel von „dem Butter“ spreche, könnte das ein Hinweis darauf sein, dass ich in Bayern aufgewachsen bin, usw. Ich könnte all diese Dinge aber auch nur vortäuschen, um die Sprachbullen in die Irre zu führen. Auch das, also die Plausibilität meines Fehlerprofils, ist Teil einer solchen Analyse. Ähnlich untersucht auch die Stilanalyse Eigenheiten meines Schreibstils. Was für Begriffe verwende ich, weist mein Satzbau spezifische Muster auf, gibt es wiederholt auftretende Begriffskonstellationen, die vielleicht sogar in verschiedenen Texten auftauchen, usw. Ich denke jede*r, die*der sich seine*ihre Texte genauer ansieht, wird einige eigene stilistische Charakteristika erkennen.

Solche qualitativen Analysen dienen vor allem dem Profiling der Verfasser*innen. Zwar können auf diese Art und Weise sicher auch unterschiedliche Texte einander zugeordnet werden, aber der eigentliche Wert solcher Analysen liegt darin, Dinge wie, Alter, „Bildungsgrad“, „Szenezugehörigkeit“, regionale Herkunft, ja manchmal vielleicht sogar Hinweise auf Berufstätigkeiten/Ausbildung, usw. bestimmen zu können. Auch Versuche, Dinge wie Geschlecht zu bestimmen, sind bekannt, scheinen aber in der Regel nicht ganz so einfach zu sein.

Demgegenüber gibt es auch eher quantitative und statistische Analysen, die von Worthäufigkeiten über Wortkonstellationen bis hin zur syntaktischen Satzstruktur alle quantitativ messbaren Sprachcharakteristika untersuchen. Diese unter dem Begriff Stilometrie geführten Verfahren sind teilweise sehr umstritten, weil nicht genau gesagt werden kann, was mit ihnen eigentlich gemessen wird/werden soll, liefern gerade in Kombination mit Ansätzen des Machine Learnings aber zum Teil erstaunliche Ergebnisse. Ich denke, dass diese Ansätze daher vor allem dazu genutzt werden dürften, verschiedene Texte nach ihren Ähnlichkeiten zu clustern.

Der klare Vorteil solcher quantitativen Analysen ist, dass diese massenhaft durchgeführt werden können. Sämtliche digital verfügbaren oder digitalisierbaren Texte lassen sich so analysieren. Vom Posting in sozialen Medien bis hin zum Buch können mit diesen Verfahren Texte erfasst werden. Zwar ist der Erfolg dieser Verfahren derzeit noch relativ bescheiden und vielfach hat sich herausgestellt, dass angeblich ähnliche Texte sich häufig mehr in ihrer Gattung geähnelt haben, als in ihrer Autor*innenschaft, aber wenn man davon ausgeht, dass individuelle Schreibstile durchaus quantitative Muster hinterlassen könnten, so heißt das im Umkehrschluss, dass wenn diese Muster erst einmal bekannt sind, eine massenhafte Zuordnung von Texten zu Autor*innen möglich sein wird.

Und was nun?

Es gab und gibt natürlich verschiedene Lösungsansätze mit diesem Wissen umzugehen und vermutlich kann man von keinem sagen, er sei besser oder schlechter als ein anderer. Wer ohnehin keine Communiqués verfasst, die*der geht diesem Problem großteils aus dem Weg, ist aber insoweit trotzdem von dem Problem betroffen, dass Beteiligungen an Publikationen und Urheberschaften von anderen Texten auf gleiche Art und Weise ermittelt werden können. Wer Texte vor Veröffentlichung verfremdet, etwa indem mehrere Personen nacheinander Passagen daraus neu- und umformulieren, etc. läuft Gefahr, bei wiederholt ähnlichen Konstellationen ebenfalls verwertbare sprachliche und stilistische Charakteristika herauszubilden oder auch daran zu scheitern, Charakteristika erfolgreich zu verschleiern. Wer meint, er*sie könne auf das Ganze scheißen, weil ohnehin keine Textproben von einer*m vorliegen oder auch, weil er*sie überzeugt ist, dass die juristische Beweiskraft der Autorenerkennung zu wacklig ist, die*der riskiert, dass in Zukunft doch irgendwie Textproben von einer*inem verfügbar sein könnten (etwa weil sie*er erfolgreich einer Autor*innenschaft überführt wird) oder sich die juristische Würdigung des Verfahrens ändert. Wer darauf vertraut, dass die Technologie (noch) nicht gut genug ist, kann durch zukünftige Entwicklungen überrascht werden. Wer technische Lösungen nutzt, um seine*ihre Autor*innenschaft zu verfremden läuft Gefahr, dabei neue Charakteristika und Spuren zu hinterlassen und zudem schlecht geschriebene Communiqués zu produzieren, die ohnehin keine*r lesen will. Wer sowieso nie irgendwelche Texte schreibt, die*der schreibt eben keine Texte.

Also tue, was immer dir am meisten zusagt, nur tue es ab nun – sofern du das nicht ohnehin schon

tatest – eben mit dem Wissen um diese Spuren und dem mulmigen Gefühl im Bauch, das schon so manch eine*n im richtigen Moment davor bewahrt haben soll, einen leichtfertigen Fehler zu begehen.

“Ein kleiner Überblick über Methoden der modernen Forensischen Linguistik zur Autorschaftsbestimmung. Der folgende Artikel versucht aus einer nicht fachlichen Perspektive einen Überblick über behördlich angewandte Methoden der Autorschaftsbestimmung bei Bekenner*in-schreiben und anderen Texten zu geben und eine entsprechende Einordnung vorzunehmen.”



Layout vom Counter-Surveillance Resource Center (csrc.link)