

»Der Sherlock Holmes kommt manchmal durch«

Der Abhörskandal Prism zeigt, wie einfach es heute ist, aus Millionen von Texten bestimmte Inhalte herauszufiltern. Mit ähnlichen Methoden geht der Sprachwissenschaftler Joachim Scharloth (TU Dresden) auf Verbrecherjagd.

ROBERT BERLIN sprach mit ihm über Sherlock Holmes, sprachliche Fingerabdrücke und Anonymität im Netz.



Videoüberwachung

Herr Prof. Dr. Scharloth, welche Rolle spielt die Sprachwis- dort gearbeitet? Welche

senschaft im digital

Bei jeder Google-A
nologie, die das Su
waltigen Datenme
die Sprachwissens
ben.

Sie forschen zur Fo
chen Sie genau?

Ein Teilgebiet der F
von anonymen Text
hungen. Dabei geht
schaft.

Wie kamen Sie auf
se für die Verbrecherjagd zu nutzen?

Prism und andere Mitschneideskandale zeigen leider, dass im Bereich der Sicherheitsinformatik in den letzten Jahren viel passiert ist. Diese Dienste nutzen schriftliche Quellen, um die Anonymität im Netz auszuhöhlen. Vor fünf Jahren fragte ich mich: Was passiert dort? Auf welcher Basis wird

JOACHIM SCHARLOTH

Joachim Scharloth, Jahrgang 1972, ist seit 2012 Professor für Angewandte Linguistik an der Technischen Universität Dresden. Er gilt als einer der wenigen Experten in Deutschland, die forensisch-linguistische Analysen mithilfe von Computersoftware durchführen. 2008 gründete er mit Noah Bubenhofer die Forschergruppe semtracks, die sich auf die automatische Analyse von Texten spezialisiert hat. Außerdem betreibt er das Blog www.security-informatics.de.

Scharloth studierte von 1992 bis 1998 Germanistik, Politische Wissenschaft und Philosophie an den Universitäten in Mainz und Heidelberg. 2010 habilitierte er sich über die Kommunikationsgeschichte der 1968er-Protestbewegung. Gast und Vertretungsprofessuren führten ihn nach Freiburg und Zürich, von 2010 bis 2012 war er Professor an der Dokkyo Universität in Tokyo.

aus
ezo-



ssen-
zier.
mes

anchmal
lem möchte

braucht einen kritischen Umgang mit den
sicherheitsdienste. Wir dürfen nicht alles

s die Leute weismachen wollen, die aus po-
litischen Gründen das Netz screenen. Vieles von dem, was

behauptet wird, funktioniert nämlich so nicht. Deshalb ist die Gefahr groß, dass Unschuldige ins Visier geraten.

Germanisten wie Sie arbeiten an der Schnittstelle zwischen Informatik und Linguistik. Wie viele Wissenschaft-



Wer gibt es auf diesem Gebiet?

Das sind in Deutschland nur sehr wenige, vielleicht zehn bis 15. Es gibt dazu keinen Aufbaustudiengang und keinen wissenschaftlichen Diskurs. Die forensische Linguistik ist als Disziplin der Computerlinguistik hier noch nicht etabliert. Das Feld wird vor allem von Informatikern bearbeitet – was grundsätzlich nicht schlecht ist. Allerdings fehlt oft ein differenziertes Wissen, wie Sprache funktioniert. Die Anzahl von Buchstabenkombinationen oder die Wortlänge sind Kategorien, die nur sehr bedingt etwas über den Sprachstil aussagen. So etwas führt dann schnell zu Missverständnissen.

Sie spielen auf den Sprachlichen Fingerabdruck an: Die Vorstellung, dass jemand anhand seiner Sprache eindeutig identifiziert werden kann.

Aus linguistischer Perspektive kann es das nicht geben. Der fundamentalste Unterschied ist: Meine Sprache ist immer sozial. Ich passe sie dem Gegenüber und der Situation an. Drohbriefe schreiben die meisten Menschen so, wie sie es einmal im Tatort gesehen haben. Mein Fingerabdruck aber ist immer derselbe, er ist unveränderbar. Wenn Sicherheits-

behörden sagen, sie können jemanden allein aufgrund der Sprache eindeutig identifizieren, ist das Unsinn.

Warum hält sich diese Vorstellung trotzdem hartnäckig, gerade in den USA?

Viele Sicherheitsinformatiker sind zu optimistisch und unkritisch, was die Aussagekraft ihrer Analysen angeht. Sie sehen Sprache als Tatwerkzeug, da gehen die Phantasien ganz weit. Andererseits haben die USA von staatlicher Seite auch viel Forschungsgeld in diesen Bereich investiert. Und dann besteht ein gewisser Druck, dass die Ergebnisse auch entsprechend ausfallen.

Sie selbst erstellen neben Ihrer Arbeit an der TU Dresden ebenfalls forensisch-linguistische Gutachten. Treffen Sie



am Ende die Entscheidung: Der war es?

Nein, die Ergebnisse sind nie endgültig. Sie sind Indizien. Wir heben oder senken nicht den Daumen, sondern liefern eine Skala: Mit dieser Wahrscheinlichkeit ist diese Person Verfasser des Textes. Es muss noch viel mehr geben, damit ein Täter überführt wird.

Was genau wird in den Texten untersucht?

Insgesamt gibt es bis zu 130 Merkmale, nach denen ich die Texte analysiere. Beispielsweise wie oft welche Nebensätze vorkommen, auch die Anzahl von Adverbien und Nominalkonstruktionen. Ich vergleiche zunächst die Schriften miteinander, von denen der Verfasser bekannt ist. Schließlich kommt der anonyme Text dazu und meine Software errechnet die Ähnlichkeit der Sprache mit den Vergleichstexten.

Also macht der Computer einen Großteil der Arbeit?

Nein, es ist nicht so, dass ich nur auf einen Knopf drücke. Wir können uns nicht verrechnen, aber wir müssen den Computer mit den richtigen Daten füttern und aus den Ergebnissen die richtigen Schlüsse ziehen – und sie relativieren.

Wer sind Ihre Auftraggeber?

Das sind Leute, die in der Öffentlichkeit stehen und in der Gesellschaft eine gewisse Rolle spielen. Sie haben einen Drohbrief erhalten oder werden erpresst. Auch bei anonymer Rufschädigung im Netz werden solche Gutachten angefordert. Ich nehme aber nur interessante Fälle an, an denen ich neue Methoden ausprobieren kann.

Haben Ihre Gutachten schon einmal einen Gerichtsprozess beeinflusst?

Nein. Meist kommt es bei diesen Streitfällen nicht zum Verfahren, sondern die Sache wird intern geklärt.

Wie lange dauert so eine Analyse?

Das hängt ganz von den Voraussetzungen ab. Wie viele Vergleichstexte habe ich? Je weniger, desto schwieriger wird es. Auch wenn ich ein Gedicht mit einem Brief vergleichen soll, ist dies aufwendiger als bei zwei Briefen. 20 Stunden brauche ich mindestens für die Analyse. Die Gutachten umfassen meist mehr als 30 Seiten. Manchmal stelle ich aber auch schnell fest, dass der Verdächtige nicht der Verfasser sein kann, weil die sprachlichen Unterschiede zu groß sind.

Wonach wird entschieden, wer verdächtig ist?

Der Kreis der Verdächtigen wird durch den Auftraggeber bestimmt. Wer hatte Zugang zu bestimmten Informationen, die im Text vorkommen? Gibt es persönliche Feinde? Als Linguisten sind wir immer von der kriminalistischen Vorarbeit abhängig. Wenn der Verfasser des Textes nicht unter den Verdächtigen ist, können wir auch nichts machen. Unser Verfahren ist nicht objektiv, sondern davon abhängig, welche Texte wir als Basis verwenden.

Wie oft liegen solche Gutachten richtig?

Schwer zu sagen: Raimund Drommel, ein Pionier der Forensischen Linguistik, sagt, er habe sich noch nie geirrt – und

das bei mehr als 700 Gutachten.

Kann man als Erpresser nicht seinen wahren Stil verschleiern?

Doch, das probieren viele auch, allerdings verändern sie meist nur einige Merkmale. Wenn man diese nach und nach ausschließt und sich zwei Texte dann sehr schnell ähnlich werden, ist es wahrscheinlich, dass sie von ein und demselben Verfasser stammen.

Durch die Abhörskandale sehen viele die Idee vom Gläsernen Menschen bestätigt. Sollten wir uns im Netz besser einen anderen Sprachstil angewöhnen?

Das wäre die falsche Reaktion. Damit würden wir unsere Freiheit einschränken. Ich sehe es als Aufgabe des Staates,



die Anonymität im Netz zu wahren. Momentan gibt es noch sehr viele Unklarheiten: Ist es erlaubt, Datenbanken von Personen zu erstellen mit den Texten, die öffentlich zugänglich sind? Darf eine Institution zum Beispiel Foren-Einträge eines Nutzers kopieren und speichern? Ich finde: Autorenprofile sollten nur bei dringendem Verdacht erstellt werden dürfen. Das muss Bestandteil des Datenschutzes werden. ■

