

Uwe Topper

Eine mathematisch-statistische Methode zur Feststellung der Autorschaft literarischer Erzeugnisse der Vergangenheit

Rezension eines Kapitels in: Fomenko, Anatoly T. (2005): „History: Fiction or Science? Chronology Vol. 2“ (Delamere, Paris etc.)

In diesem hochinteressanten und sehr eigenwilligen zweiten Band der neuen englischen Buchreihe von Fomenko befindet sich als Annex 3 (S. 425-444) ein Beitrag der Eltern von Anatoly T. Fomenko: Timofei G. und Valentina P. Fomenko, die von 1974 bis 1981 in Moskau ein Forschungsprojekt durchführten, das erstaunliche Ergebnisse gebracht hat, deren Anwendung für unsere chronologiekritische Analyse in hohem Grade nützlich sein könnte.

Nach einer gekürzten Veröffentlichung 1983 im Institut für sowjetische Geschichte an der Akademie der UdSSR wurde die Gesamtarbeit 1996 herausgegeben.

Es geht um die Erkennung der Autorschaft eines beliebigen Textes, der genügend Einzelelemente (Worte, Sätze) enthält, die seine individuellen Eigenschaften untersuchbar machen. Wenn von einem Autor ein größeres Werk oder mehrere Werke gleicher Gattung vorliegen, müsste es ein Leichtes sein, die häufig wiederkehrenden Begriffe, Wendungen und Eigenarten im Stil statistisch zu erfassen und anhand dieser Merkmale ein weiteres Werk desselben Autors, dessen Urheberschaft unbekannt ist oder absichtlich verschwiegen wird, zu erkennen. Die staatliche Nutzenanwendung in den achtziger Jahren des 20. Jahrhunderts und die damit verbundene Förderung dieser Forschung in der Sowjetunion scheint mir offensichtlich, wenn man die unüberschaubare Menge an Untergrundliteratur (Samizdat) in Betracht zieht, die damals im Umlauf war. Die Nutzenanwendung für uns heute sehe ich darin, dass man mithilfe dieser Methode einen angeblich antiken Autor mit einem bekannten Renaissance-Schriftsteller identifizieren könnte, wenn das Verfahren eindeutig ist und eine entsprechend große auswertbare Datenmenge vorliegt.

Nun ist es ja ein bekannter Trick, dass ein Autor, der seine Person verstecken will, in einem Werk, das er zusätzlich zu seinen schon bekannten Schriften herausgeben möchte, seinen Stil verändert, gewisse bekannte Reizwörter vermeidet, eigentümliche Wendungen abändert usw. – das kann einen Neugierigen durchaus täuschen. Es gibt aber Sprachelemente, die nicht nur ein sehr intelligenter Pseudonymbenützer übersieht, sondern die auch fast unmöglich bewusst geändert werden können, nämlich kleine Wörtchen wie Präpositionen, Bindewörter, Umstandswörter usw. Auf dieser Erkenntnis, die sich aus den jahrelangen Untersuchungen des Ehepaars Fomenko herausfilterten, beruht die Methode, die die Fomenkos erfolgreich auf ein umstrit-

tenes Literaturwerk, den „Stillen Don“, anwandten. Die Autorschaft dieses Werkes wurde gemeinhin dem bekannten Dichter Scholochow zugeschrieben, aber es wurde auch ebenso lange Zeit heftig angezweifelt, dass es von Scholochow stamme. Aus der vorgelegten mathematischen Analyse der Werke von Scholochow im Vergleich zum Prosatext „Der stille Don“ können Fomenkos mit Sicherheit schließen, dass dieser von einem anderen Dichter geschrieben sein muss.

Schauen wir uns die Methode an.

Als Vorläufer der Arbeitsweise nennen die beiden Autoren zunächst W. Fuchs (London 1955 und Stuttgart 1968), der Silbenzahl und Satzlänge als eigentümliche Werte eines Textes in Betracht zog. Demgegenüber hatte schon A. A. Markov 1916 festgestellt, dass natürlicherweise viele der untersuchten Worthäufigkeiten und Verwendungsarten „sich um einen gemeinsamen Wert herum gruppieren müssen aus sprachimmanenten Regeln“, was die zweifelsfreie Zuordnung der Autorschaft erschwert.

Die Fomenkos fordern daher, dass als unzweifelhafte Eigenschaften einerseits solche zählen müssen, die einem Autor oder einer kleinen Gruppe gleicher Autoren gemeinsam sind, und dass sie andererseits bei anderen Autoren nicht in gleicher Weise wiederkehren dürfen.

Als unbrauchbar scheiden sofort alle jene Wörter und Satzbildungen aus, die ein Dichter bewusst wählt, um seinem Stil eine besondere Note zu verleihen. Nur unbewusste Charakteristika kommen für die Beurteilung in Betracht.

Für die Untersuchung müssen zwecks Abgrenzung außerdem eine große Anzahl von Werken verschiedener Autoren oder Gruppen zur Verfügung stehen.

Eine völlige Auswertung aller literarischen Texte, die in Frage kommen, war seinerzeit eine arbeitsmäßige Unmöglichkeit, deshalb musste ferner noch eine statistisch verwertbare Methode der Beispielauswahl angewandt werden, die möglichst gleichbleibend durchführbar sei. Mit den damals noch nicht so weit entwickelten technischen Fähigkeiten computergestützter Analysen ist diese Vorsichtsmaßnahme verständlich, dürfte aber heute eine weit geringere Rolle spielen. Natürlich stehen einem Autor heute, wenn er sich „verstecken“ will, dieselben Computerdienste zur Verfügung, weshalb sich wie immer mit dem technischen Fortschritt auch die Notausgänge und Schlupflöcher in gleichem Maße vervielfältigen, wie man glaubt, sie einschränken zu können. Da ich nur im

Sinn habe, die Methode für unsere geschichtliche Analyse, besonders bezüglich der Renaissance-Autoren, anwendbar zu machen, können wir das Nutzungsfeld „Samizdat“ hier getrost übergehen.

Die Fomenkos wandten ihre Untersuchung zwecks Nutzbarmachung der Methode für einen erstaunlich weiten Bereich von russischen Literaturwerken an, beginnend im 18. Jahrhundert mit sechs Autoren (von Tschulkow bis Krylow), im 19. Jahrhundert mit neun Dichtern von Gogol bis Tolstoi und im 20. Jahrhundert mit weiteren acht von Gorky bis Scholochow. Eine der wichtigen Erkenntnisse war, dass ab einer ausreichenden Menge von Wörtern, hier 16.000, eine signifikante Häufigkeit ihrer Verwendung als Kurve erkennbar gemacht werden konnte. Aus einer Liste von neun untersuchten Eigenheiten ergab sich schließlich, dass nur eine Gruppe tatsächlich als Invariante nutzbar zu machen war, während die anderen höchstens in ihrer Kombination gewisse Aussagen zuließen. Diese Invariante betrifft die „kleinen“ Wörtchen, besonders die Häufigkeit der im Russischen sehr wichtigen „in“ und „nicht“.

Tatsächlich zeigen die beiden Autoren mit mathematischer Sicherheit, dass Scholochow nicht den „Stillen Don“ geschrieben haben kann, und auch, dass es möglicherweise sein Zeitgenosse Fjodor D. Kryukow gewesen sein könnte, der damals in Russland noch weitgehend unbekannt war und von dem nicht ausreichend Textmaterial zwecks Analyse zur Verfügung stand.

Wer sich für die mathematische Seite der Untersuchung interessiert, sollte diesen Beitrag in Fomenkos Buch unbedingt lesen. Die beigegebenen Tafeln und Kurven sprechen für sich. Ich möchte nämlich anregen, dass sich im Laufe der Zeit – mit dem Vorhandensein einer großen Zahl elektronisch erfasster und zugänglich gemachter Texte der Renaissance – junge technisch begabte Geschichtsanalytiker in dieser Weise an der Arbeit beteiligen und so die Verbindungslinien zwischen den „Herausgebern“ der antiken Werke und den tatsächlichen Autoren aufzeigen. Unsere bisherige Betrachtungsweise der Renaissance-Werke war weitgehend künstlerisch bestimmt, vertraute also auf ein gewisses Empfinden für Stil und geistige Voraussetzung einer Person, was nur ansatzweise Ergebnisse erzielen konnte, während eine mathematisch-naturwissenschaftliche Untersuchung eine weit größere Zahl von Menschen von der Richtigkeit unserer Vermutungen überzeugen könnte. ■