

Carole E. Chaski

FAKERY, FORGERY, AND FORENSIC COMPUTATIONAL LINGUISTICS

The question of authorship has a long history in literary and historical studies. Some well-known examples of authorship disputes include the Epistles of the New Testament,¹ Shakespeare's plays,² and *The Federalist Letters*.³ Authorship identification can also play a role in the forensic setting, where the authorship of a document has legal consequences. This chapter focuses on fakery, forgery, and a tool for linguistic evidence in the forensic setting. This tool, ALIAS: Automated Linguistic Identification & Assessment System, was developed specifically for authorship identification, later expanded to include other issues such as suicide note authentication, and has been used in both literary and historical disputes as well as legal cases.

1. Overview of ALIAS:

Automated Linguistic Identification & Assessment System

ALIAS is a web-accessed system for providing sophisticated, doctoral-level linguistic analysis to non-linguists, such as law enforcement, human resources professionals or scholars in the social sciences and adjacent fields. ALIAS is

- 1 The Epistles dispute mainly concerns the Letter to the Hebrews, with different views on whether Paul or someone else authored it, a dispute that is documented from Eusebius's Ecclesiastical History (312–324 AD).
- 2 For Edward de Vere's view on the matter, see <https://shakespeareoxfordfellowship.org/discover-shakespeare/>, accessed: 11 July 2024. For the Shakespeare authorship theory, see <https://shakespeareauthorship.com/howdowe.html>, accessed: 11 July 2024.
- 3 For a historical analysis, see Douglass Adair: *The Disputed Federalist Papers*, in: *Fame and the Founding Fathers*, ed. by idem. Indianapolis 1974. This work is based on much earlier work by Adair. For a statistical analysis (which generally follows Adair), see https://en.wikipedia.org/wiki/Frederick_Mosteller, accessed: 11 July 2024; David L. Wallace, 2012: *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. https://books.google.de/books?id=LJXaBwAAQBAJ&redir_esc=y, accessed: 11 July 2024. Updated 2nd edition, originally published as *Frederick Mosteller*; David L. Wallace: *Inference in an Authorship Problem*, in: *Journal of the American Statistical Association* 58.302 (June 1963), pp. 275–309 (<https://www.jstor.org/stable/2283270>, accessed: 11 July 2024).

multi-functional, because it applies forensic methods to the four main areas where language might serve as evidence:

- (1) identification of authorship, language and writing system;
- (2) authenticating text such as suicide notes, threats, or predatory chats;
- (3) textual relationships and similarity; and
- (4) linguistic profiling.

ALIAS is *multi-lingual*, with all algorithms running in English, and with at least one algorithm from each of the four main areas running in many other or all languages. It has multiple *methods of data entry*, including dictating, copy-pasting, uploading text files and email archives, unzipping files, transcribing speech, transcribing handwriting, ingesting social media, and retrieving machine-generated texts from current suppliers. The system has specific processes for handling microtexts (such as tweets) and macrotexts (such as novels and deposition transcripts).

ALIAS consists of three components for the three main functions of text analysis (TATTTLER), forensic text analysis (FORTAN), and accounting (ATLAS). ALIAS TATTTLER provides a full text analysis toolkit, for example lemmatizing, lexicon, n-gramming, syntax, semantics, discourse, sociolinguistics, phonology, punctuation, c-gramming, and scripts (Roman, Cyrillic, and Arabic). These kinds of analysis can be combined and tested statistically to create methods for forensic linguistic evidence which are then incorporated into ALIAS FORTAN. Currently, ALIAS FORTAN provides forensic text analysis methods such as SynAID (syntactic author identification), SNARE (suicide note authentication), HANDLER (handwriting identification support), and more than a dozen others. ALIAS ATLAS provides time tracking for invoicing, and client and case management.

2. Fakery in the Forensic Setting

Faked documents are *completely invented* and presented as legitimate. Sometimes, faked documents are anonymous or presented with a name of a person who does not exist. In other cases, a faked document is presented with the name of a real person who denies authorship.

On 29 April 1992, a healthy 28-year-old, Michael Hunter, died in his sleep. Blood analysis showed a lethal injection of lidocaine, Benadryl, and Vistaril. Independently, these drugs are benign, but their mixture is deadly. Hunter

had a prestigious college degree and a good job at an educational technology company. He was close to his parents, a devout Christian, and sang in his church choir. He was well-liked, a friendly, witty guy who had been involved with one partner for a long time.

Hunter lived with two roommates. One of them, Gary Walston, his longtime partner, was out of town when Hunter died. The other roommate, Joseph Mannino, had reported the death. A fourth-year medical student, Mannino was only weeks away from completing his medical degree. Mannino was interviewed several times during the investigation of the death. At first, he admitted giving antihistamines to Hunter for a migraine headache. Later, after an autopsy report showed the presence of lidocaine, Mannino denied injecting Hunter with the drug, which can induce central nervous system collapse at specific dosages. He declared that Hunter must have injected the drug himself. Finally, Mannino produced computer disks containing two suicide notes from Hunter, one addressed to Hunter's parents and one to Gary, the roommate with which Hunter had lived for quite a while.

Detective W. Allison Blackman of the Raleigh, North Carolina, Police Department contacted me about the possibility of determining the authorship of the suicide notes on the basis of the language used. Detective Blackman gathered almost 10,000 running words of documents spontaneously written by Hunter, Mannino, and Walston. These included student essays, personal letters, and notes. All of the documents, both the ones with known authorship and the suicide notes of questioned authorship, were analyzed in the same way using the original version of ALIAS SynAID. Syntactic author identification analysis refers to the parsing of sentences into phrases and phrases into words; phrase types can then be categorized into various classes. The syntactic patterns are then counted, with this quantification used in statistical analysis. Likewise, vocabulary and punctuation patterns can be classified, counted, and statistically tested.

SynAID statistically differentiated the three known authors – and the possible candidate authors of the suicide notes – Hunter, Walston, and Mannino. Detective Blackman wanted to know which of the three men who had access to the home computer could have authored the suicide notes. Further, the statistical testing showed that the suicide notes were not authored by Hunter (where the probability of his authorship was about 1 in 10,000), nor by Walston (where the probability of his authorship was about 1 in 100), but most likely authored by Mannino, for whom there was no statistical difference between his known documents and the suicide notes. SynAID statistically excluded Hunter as the author of the notes, Walston as the author of the notes, but included Mannino as the author of the notes. Because each

questioned document is tested individually, the SynAID method could have associated one of the notes to Hunter and the other to Mannino, or any of the possible combinations; instead, SynAID associated both notes to Mannino, and neither note to Hunter or Walston.

On the basis of the linguistic analysis, which provided the legally required probable cause, Mannino was charged with first-degree homicide. Through his attorney, Mannino was willing to accept a plea deal in which he confessed to second-degree homicide, but the prosecutor rejected this offer and trial proceeded with the first-degree homicide charge. On the witness stand, Mannino admitted to having supplied Hunter with lidocaine. He admitted that he had authored the fake suicide notes in order to draw the police attention away from himself.

The medical examiner's toxicology report showed the presence of three drugs, including lidocaine. The combination of lidocaine with Benydril and Vistaril can cause heart failure within two minutes, but this can depend on the amount of lidocaine being administered. Mannino's claim that Hunter injected himself with lidocaine was not credible, but the level of lidocaine became a crucial issue. The bulk of the scientific literature specified lethality at one level, while one article specified another. Based on this one ambiguous test result in the scientific literature and given the level of lidocaine in Hunter's blood, the defense argued that the amount of lidocaine might not have been lethal, and that perhaps Mannino had only ›accidentally‹ killed Hunter. Mannino was then convicted of the lesser charge of involuntary manslaughter (rather than first-degree homicide). In the sentencing phase of the trial, Judge Stephens said, »It's terrible that Michael Hunter died. It's terrible that the defendant unlawfully caused his death. But to give the impression that Michael Hunter took his own life, I find that extremely aggravating in this case«.4 On 29 July 1994, Mannino was sentenced to seven years in prison.

3. Forgery in the Forensic Setting

Forged documents are *submitted under a false name*, authored by someone who is not identified publicly as an author, and the public authors must have to some degree colluded with the forger.

For almost two decades, shareholders of the Yukos Oil Company and the Russian Federation have been in a dispute regarding the legality of Yukos' expropriation by Russia on grounds of tax evasion, which led to the bank-

4 The News and Observer, Raleigh, 28 July 1994.

ruptcy of the company.⁵ Trials or filings related to this case have taken place in the Netherlands, France, Belgium, the United States, the United Kingdom, and India and the Yukos shareholders are planning even more litigation. This dispute was originally negotiated in an international setting in 2007 at the Permanent Court of Arbitration in The Hague, with a mutually-agreed upon Tribunal (Swiss lawyer Charles Poncet, American jurist and international judge Stephen M. Schwebel, and Canadian lawyer and international mediator L. Yves Fortier) who employed a secretary for aid, Martin J. Valasek. In the Court's original decision, the Yukos Oil Company shareholders won a 54-billion-dollar arbitration award against the Russian Federation, which was promptly appealed on several grounds, including jurisdiction and fraud.

On 20 April 2016, on appeal, the decision in favor of the shareholders was overturned by the District Court of the Hague. Again, this decision was appealed, and in February 2020, the Court of Appeal of the Hague reinstated the 54-billion-dollar award to the Yukos shareholders. But on 5 November 2021, the Supreme Court of the Netherlands also struck down the award and referred the case back to the Amsterdam Court of Appeal, this time with a focus on fraud. The Russian Federation based its appeal on an allegation of fraud by the Yukos shareholders. In February 2024, the Amsterdam Court of Appeals »rejected a final ground of appeal filed by Russia alleging fraud by former shareholders. The court said in a written statement that Russia made the claim too late in the proceedings. It added that even if Russia had raised the alleged fraud at an earlier phase of the drawn-out case it would not have altered the outcome.«⁶

One of the issues which the Russian Federation raised in this case is the authorship of the award granting the 54-billion-dollar fine against the Russian Federation. The Russian Federation argued that the Tribunal did not author the award, as expected by arbitrational rules and customs. On this point, in commentary of the Yukos case, New York University law professor Rob Howse quotes from an expert opinion provided by Columbia University law professor George Berman in a proceeding at the District of Columbia Circuit, Washington, DC, United States. Professor Bermann opined:

5 Yukos Universal Limited (Isle of Man) v The Russian Federation, PCA Case No. 2005-04/AA227 (<https://pca-cpa.org/en/cases/61/>, accessed: 11 July 2024).

6 The Associated Press: Dutch court rejects final argument in legal battle over former Russia oil giant Yukos, see: <https://www.bilaterals.org/?dutch-court-rejects-final-argument>, accessed: 24 April 2024.

It is generally viewed as impermissible for arbitral secretaries to produce even a preliminary draft of substantive portions of an award. This is so, irrespective of the degree of care a tribunal brings to its subsequent review of the draft.

Further, Professor Berman explains that there:

exists in international arbitration a powerful consensus that for arbitral secretaries to draft substantive portions of the awards is off limits. The consensus in this respect is especially clear and overwhelming.⁷

Russian Federation attorneys argued that, in contrast to this overwhelming consensus, the award was substantially authored by the secretary, based in part on the inordinate amount of time that Mr Valasek clocked in the case. The Russian Federation attorneys contacted ALIAS Technology to determine if their hypothesis about the awards' authorship was correct. Fortunately, the attorneys did not tell me about the background of the case or the anomalous amount of time that Mr. Valesek spent as a secretary rather than an arbitrator.

In fact, applying ALIAS SynAID to the known writings of the three arbitrators (Poncet, Schwebel, and Fortier) and the secretary Valasek attained a statistical model that differentiated the authors' known documents at close to 100% accuracy. Applying that statistical model to the Awards, 70% of the Awards was assigned to the secretary at 95% or higher likelihood of his authorship. After this analysis was completed, the attorneys contacted computational linguist Walter Daelemans to review the statistical analysis; he concurred with the SynAID conclusion.

During the 2016 trial in the District Court of the Hague, arbitrator Fortier for whom the secretary Valasek worked, testified that »Mr. Valesek did not write the tribunal's *reasoning and conclusions* of the Yukos awards«. ⁸ In an insightful and prescient comment written in 2017, Professor Howse noted that »Arbitrator Fortier's statement avoided a precise answer to the question posed to him, or its substance«⁹ because arbitrator Fortier provided the loophole that was eventually exploited in another ruling four years later. In a 2021 ruling from the Supreme Court in the Netherlands, the issue of authorship was addressed.

7 Rob Howse: The fourth man: an intriguing sub-plot in the Yukos arbitration, in: International Economic Law and Policy Blog, 29 March 2017 (<https://www.bilaterals.org/?the-fourth-man-an-intriguing-sub>, accessed: 11 July 2024).

8 Ibid.

9 Ibid.

2.28. In paras. 6.6.1-6.6.15, the Court of Appeal addressed the Russian Federation's argument that the Yukos Awards should be set aside due to the disproportionate involvement of Martin Valasek, the assistant to the Tribunal, in their drafting (para. 6.6.1). According to the Russian Federation, this involvement violated the principle that arbitrators must personally perform the task assigned to them, meaning that the Tribunal had failed to comply with its mandate (Article 1065(1)(c) DCCP). In addition, Valasek's involvement is said to mean that there effectively had been a »fourth arbitrator«, meaning that the Tribunal's composition was inconsistent with the applicable rules (Article 1065(1)(d) DCCP).

2.29 The Court of Appeal rejected this argument. *Although the Court of Appeal presumed that Valasek had made significant contributions to the drafting of parts of the text (para. 6.6.5), this did not mean that he had independently taken decisions which are part of the essential task of the arbitrators (para. 6.6.10).* The fact that the Final Awards were signed by the three arbitrators implies, that it was they who rendered them, which means there was not an even number of arbitrators (para. 6.6.13). If the Russian Federation's statement that Valasek was only introduced as an assistant and contact person is assumed, by way of presumption, to be correct, it can be concluded that the Tribunal failed to fully inform the parties on this point of the nature of Valasek's work. However, under the circumstances, this does not constitute such a serious violation of the mandate that it should lead to the setting aside of the arbitral awards (para. 6.6.14.2).¹⁰ (italics added, CEC)¹¹

Arbitration specialists Omar Puertas and Borja Álvarez comment on this:

The Court of Appeal's decision did not hinge on weighing the linguistics' expert evidence, but rather on the question of whether proof had been furnished to demonstrate Valasek's actual participation in *the decision-making process*. In illustrative terms, the Court of Appeal reasoned that »it is up to the expertise of Chaski and Daelemans to analyze, using scientific methods, which author most likely wrote a certain text, but not to determine whether that author wrote that

10 Opinion of the Attorney-General at the Supreme Court of the Netherlands, Friday, 23 April 2021 (<https://jusmundi.com/en/document/decision/en-yukos-universal-limited-isle-of-man-v-the-russian-federation-conclusion-of-the-attorney-general-at-the-supreme-court-of-the-netherlands-friday-23rd-april-2021>, accessed: 11 July 2024).

11 See also Omar Puertas and Borja Álvarez: The Yukos Appeal Decision on the Role of Arbitral Tribunal's Secretaries (<https://www.ibanet.org/article/B55CB7F1-01C6-4BDF-9383-90F567C17147>, accessed: 11 July 2024).

text on his own authority or on the instructions and under the responsibility of someone else (para 6.6.6). (Italics added, CEC)

In summary, in this highly complex civil case involving international bilateral treaties and multiple jurisdictions, the Courts have accepted the scientific evidence of authorship of a substantial part of the contested Award, using ALIAS SynAID, but the Attorney General of the Supreme Court of the Netherlands has found a way out of the consequences of the forged authorship by concluding that the authorship of the documents does not entail the decision making – »the reasoning and the conclusions« – in the documents.

4. Suicide Note Authentication

After the Nirvana rock star Kurt Cobain died on 5 April 1994, conspiracy theorists (fostered by pseudo-documentaries and a private detective's theory) speculated that Cobain's wife Courtney Love had killed Cobain – or at least that she had authored part of his suicide note. As part of an investigation for the docudrama *Soaked In Bleach* (2015), ALIAS Technology was asked to authenticate the suicide note using ALIAS SNARE, one of the methods in ALIAS FORTAN. The producers requested confirmation – based on linguistic evidence – of whether the suicide note was a fake or a forgery or authentic.

The remarkable aspect of Cobain's suicide note is that it has two distinct parts, now known as the top and the bottom. The longer top part describes his depression rooted in his disenchantment with the music industry; the shorter bottom part expresses his love for Courtney and his daughter. The top is not as easily recognized as a suicide note by laypersons who have not studied them, while the bottom is closer to stereotypical ideas about suicide notes. Conspiracy theorists claimed that Kurt had authored the top (but that it had not been intended as a suicide note) and that Courtney had authored the bottom (to turn the letter into a fake suicide note).

ALIAS Technology ran three tests. ALIAS SNARE authenticated (i) the entire note, (ii) the top part and (iii) the bottom part as suicide notes, rather than control documents. The ALIAS SNARE method was developed using almost 500 authentic suicide notes, and close to 1,000 control documents which are similar to suicide notes in specific ways. For instance, authentic suicide notes can contain elements once might expect to find for instance in love letters, trauma narratives, or even business letters. The longer the suicide note, the more likely it is that it contains such elements. Shorter

suicide notes are therefore easier to detect because they are not long enough to contain elements that we have learned to associate with love letters et cetera. SNARE differentiates between authentic suicide notes and control documents with 80% accuracy on longer documents and 86% accuracy on shorter documents.¹² Based on the ALIAS SNARE results, I explained what I had found to one of the producers who was in contact with me and was invited to appear in the movie.

I was filmed for *Soaked in Bleach*, a project which represented the views of the private detective, but disregarded the opinions of the experts consulted on the case, which unanimously agreed that Cobain's death was a suicide, as had been determined by the Seattle Police Department. For example, retired New York City Police Detective Vernon J. Geberth posted an article on his *Practical Homicide Investigation* website¹³ and Facebook page,¹⁴ stating that he »was not happy that the producers of *Soaked in Bleach* made it appear that he agreed with their homicide theory«. He stated further that he »made it quite clear that he believed that Kurt Cobain took his own life and backed up his opinion with the facts that he had obtained from the Seattle Police Department's Homicide Division coupled with his own experience with suicide cases«. ¹⁵ Likewise, when I was interviewed on the NBC affiliate House Of Mystery Radio Show after the release of *Soaked in Bleach*, I stated that »my results do not support the conspiracy theory that Courtney Love authored the bottom portion to make it look like a suicide note«. ¹⁶ My discussion of this fact was deleted from the movie since it did not fit the private detective's pet theory, and conspiracy theorists invited me to speak at their conference, until I explained that the portrayal of the forensic computational linguistics in *Soaked in Bleach* was a misrepresentation of the actual results. In actuality, what I had explained about the disputed suicide note was not included in the movie in whole, so that only one part of the facts was excerpted from my interview and cast in a way to support a conspiracy theory against Courtney

12 Carole E. Chaski and Denise Huddle D'Abre: Is This a Real Suicide Note? Authentication Using Statistical Classifiers and Computational Linguistics, in: Proceedings of the American Academy of Forensic Sciences, 2012, p. 439.

13 Quoted from Wikipedia, retrieved on 25 June 2017 (<https://www.practicalhomicide.com/Research/research.htm>, accessed: 24 April 2024).

14 Quoted from Wikipedia, retrieved on 25 June 2017 (<https://www.facebook.com/PracticalHomicideInvestigation/posts/1016939241687918/>, accessed: 24 April 2024).

15 Quoted from Wikipedia (https://en.wikipedia.org/wiki/Soaked_in_Bleach, accessed: 24 April 2024).

16 Quoted from Wikipedia, retrieved on 13 March 2021 (https://en.wikipedia.org/wiki/Soaked_in_Bleach#cite_note-24, accessed: 24 April 2024).

Love. The movie misrepresented the fact that the top note, the bottom note and the whole note were all classified by SNARE as authentic suicide notes, because SNARE is built on data that includes both longer and shorter notes.

Cobain authored two types of suicide notes – a sophisticated one and a stereotypical one. The sophisticated suicide note, the top, would not be recognized as an authentic suicide note by laypersons or private detectives who are not familiar with suicide notes. By suicide note standards, it is long, and as a longer suicide note, it contains elements typically found in love letters, trauma narrative, and business letters. The stereotypical note, the bottom, is what most laypeople think a suicide note should sound like, and it is typical of short notes. What is most interesting about Cobain's note is its structure and the parallel to his actions. He wrote a long note and he wrote a short note. Cobain killed himself in two ways, first by a lethal overdose and then a gunshot to his head. The structure of the two-part note is an eerie linguistic echo of his actions, a dual-method suicide. Even in death, Cobain's command of language echoed his choices in life.