Drafting Standards for Stylometry

Juola, Patrick
juola@mathcs.duq.edu
Duquesne University, United States of America

Byszuk, Joanna
joanna.byszuk@ijp.pan.pl
Institute of Polish Language of the Polish Academy of Sciences, Poland

Stylometry and authorship attribution [1] have matured greatly since the early research in the late 20th century. With the development of formal algorithms for document comparison [2,3], the formalization of evaluation techniques [4,5,6,7], and the development of standardized software for analysis [8,9,10], it is now easier than ever to analyze the writing style of documents. Several recent high-profile cases [11,12] demonstrate this. Indeed, courts across the world are becoming more and more willing to accept analyses of writing style as proof of authorship [13,14,15,16,17].

Unfortunately, the increased use of stylometric evidence has created opportunities for the admission and reliance upon erroneous and unreliable stylometric evidence [18,19]. Both Grant [18] and Coulthard [19], described many issues with the $50 million Yukos Arbitration Case [18,19], a still-pending (as of this writing) dispute about whether "the Russian Federation had illegally seized assets from Yukos," [19, p. 49] and, later, about whether the initial arbitration had been illicitly arrived at. Two analysts had independently arrived at conclusions that, while similar in their overall conclusion (that the arbitration had been illicitly written), contradicted each other on key aspects (such as which specific sections had been so written; indeed, of 28 separate sections, the two experts agreed on only 8.) [19]. Simple logic dictates that at least one of these conflicting reports must be incorrect and should not be relied upon.

Coulthard [19] and Juola [20] have identified several questionable assumptions in the methods applied, including

- That "there is a fixed and known number of authors
- "That the 'known' texts chosen were suitable for purpose
- "That all attributed sections are single author
- "That all sections are long enough for the chosen method of analysis
- "That variation in section length is not important
- "That no 'foreign' text remains in the sections" [19]
- And that co-authored writing can be studied as a "mixture" of individual writings. [20]

Violations of any of these assumptions are likely to produce bad evidence [21] that could lead to a miscarriage of justice (and, in the case of *Yukos*, may still). Other potential issues include the use of unreliable analysis techniques such as average word length or QSum [22] or analyst overconfidence.

The issue of reliability in forensic science has been raised before, by the US National Academy of Sciences [23] and the President's Council of Advisors on Science and Technology [24] among others. PCAST, in particular, noted the need for "validity as applied," noting that even good science can be badly used *in practice*. For example, a correctly performed statistical calculation that relies on underlying independence assumptions can produce wildly incorrect estimates of probabilities [17], or correct estimates can be grossly misrepresented as in the "prosecutor's fallacy." Even simple human error---analyzing the wrong file---can creep in. PCAST [24, p. 28] therefore called upon the scientific community to "develop and promulgate standards and guidelines to improve best practices in the forensic science community."

The proposed pre-conference workshop is open to any party with an interest in discussing such standards as they apply to stylometry. While it is impractical to establish a list of best practices in an active research community---what is best today may be improved upon tomorrow---it should at least be possible to list types of potential error and to make recommendations about ways to avoid them. By establishing guidelines and guard rails, the relevant scholarly community (and the digital humanities community contains many of the world's top experts in stylometric analysis) can provide public, written, guidance which can be relied upon by legal professionals in assessing the reliability of "expert" evidence. Indeed, under US law, the existence of "standards controlling the technique's operation" is a key factor supporting admissibility decisions. By extension, failure to follow such standards is an easily understandable reason to object to a particular report. [23,24]

This workshop is planned to take half a day. It will start with an introductory presentation about stylometry by Joanna Byszuk, move to a presentation by Patrick Juola about standards and the standards writing process [23], and then spend the bulk of the time in brainstorming and drafting ideas about important issues in stylometry affecting reliability and quality control [19,25]. It is the organizers' hope that this workshop will create a sufficiently detailed draft ("the Graz Draft"?) to enable continued work by interested participants following the conference.

Reliability in forensic science is an important social issue, even a "crisis," and one frequently addressed in the DH community over the last few years [7,13] The *Yukos* case alone involves a dispute of $50 billion that may or may not go to enrich a belligerent autocratic state; other cases have involved the potential for personal harm to individuals via potentially unjust fines, deportation, or detention. The scholars gathered at the 2023 DH conference in Graz create an opportunity to address and mitigate an ongoing problem of increasing relevance, and so to have lasting public impact.

Bibliography

Neal, Tempestt / **Kalaivani Sundararajan** / **Aneez Fatima** / **Yiming Yan** / **Yingfei Xiang** / **Damon Woodard** (2017): "Surveying stylometry techniques and applications." ACM Computing Surveys (CSuR) 50, no. 6 (2017): 1-36.

J. F. Burrows (1989): "'An ocean where each kind. . . ': Statistical analysis and some major determinants of literary style," Computers and the Humanities, vol. 23, no. 4–5, pp. 309–21.

J. F. Burrows (2002): "Delta: A measure of stylistic difference and a guide to likely
authorship," Literary and Linguistic Computing, vol. 17, pp. 267–287.

Juola, Patrick (2004): "Ad-hoc authorship attribution competition." In Proceedings of the Joint Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing, pp. 175-176.

Argamon, Shlomo / **Patrick Juola** (2011): "Overview of the international authorship identification competition at PAN-2011." In CLEF (Notebook Papers/Labs/Workshop).

Stamatatos, Efstathios / **Walter Daelemans** / **Ben Verhoeven** / **Martin Potthast** / **Benno Stein** / **Patrick Juola** / **Miguel A. Sanchez-Perez** / **Alberto Barrón-Cedeño** (2014): "Overview of the author identification task at PAN 2014." In CLEF 2014 Evaluation Labs and Workshop Working Notes Papers, Sheffield, UK, 2014, pp. 1-21.

Schöch, Christof / **van Dalen-Oskam, Karin** / **Antoniak, Maria** / **Jannidis, Fotis** / **Mimno, David** (2020): Replication and Computational Literary Studies. Digital Humanities Conference 2020 (DH2020), Ottawa, Canada. https://doi.org/10.5281/zenodo.3893428

Juola, Patrick / **John Sofko** / **Patrick Brennan** (2006): "A prototype for authorship attribution studies." Literary and Linguistic Computing 21, no. 2 (2006): 169-178.

Eder, Maciej / **Mike Kestemont** / **Jan Rybicki** (2013): "Stylometry with R: a suite of tools." In DH, pp. 487-488.

Eder, Maciej / **Jan Rybicki** / **Mike Kestemont** (2016): "Stylometry with R: a package for computational text analysis." The R Journal 8, no. 1 .

Juola, Patrick (2015): "The Rowling case: a proposed standard analytic protocol for authorship questions." Digital Scholarship in the Humanities 30, no. suppl_1 (2015): i100-i113.

Segarra, Santiago / **Mark Eisen** / **Gabriel Egan** / **Alejandro Ribeiro** (2016): "Attributing the Authorship of the" Henry VI" Plays by Word Adjacency." Shakespeare Quarterly 67, no. 2 (2016): 232-256.

Janet Ainsworth / **Patrick Juola** (2019): "Who Wrote This?: Modern Forensic Authorship Analysis as a Model for Valid Forensic Science", 96 Wash. U. L. Rev. 1161 (2019). Available at: https://openscholarship.wustl.edu/law_lawreview/vol96/iss5/10

Carole E. Chaski (2005): Who's at the Keyboard? Authorship Attribution in Digital Evidence Investigations, 4 INT'L J. DIGITAL EVID. 1.

Malcolm Coulthard (1994): Powerful Evidence for the Defence: An Exercise in Forensic Discourse Analysis, in LANGUAGE AND THE LAW 414 (John Gibbons ed.).

Malcolm Coulthard (2013): On Admissible Linguistic Evidence, 21 J. L. & POL'Y 441, 444–46.

Patrick Juola (2013): Stylometry and Immigration: A Case Study, 21 J. L. & POL'Y 287.

Grant, Tim (2022): The Idea of Progress in Forensic authorship analysis. Cambridge University Press.

Coulthard, Malcolm (2022): "Rigour and Transparency in Forensic Linguistics Casework" Keynote at IAFLL 2022, Porto, Portugal.

Juola, Patrick / **Alejandro Napolitano Jawerbaum** (2022): "Coauthored Documents Are More than Mixtures of the Styles of the Component Authors." IAFLL 2022, Porto, Portugal. Available at https://iafllporto2022.files.wordpress.com/2022/07/iafllporto2022-bookabstracts-3.pdf

Rudman, Joseph (1997): "The State of Authorship Attribution Studies: Some Problems and Solutions." Computers and the Humanities 31, no. 4 (1997): 351–65 Available at http://www.jstor.org/stable/30200436.

Holmes, David I. (1998): "The evolution of stylometry in humanities scholarship." Literary and linguistic computing 13, no. 3: 111-117.

National Research Council (2009): Strengthening forensic science in the United States: a path forward. National Academies Press.

Lander, Eric S. / **PCAST Working Group** (2016): "Forensic science in criminal courts: ensuring scientific validity of feature-comparison methods".

Juola, Patrick (2022): "Unreliable (Linguistic) Evidence and the Role of Standards." IAFLL 2022, Porto, Portugal. Available at https://iafllporto2022.files.wordpress.com/2022/07/iafllporto2022-bookabstracts-3.pdf