

CZECH LITERATURE STUDIES

PETR PLECHÁČ

# Versification and Authorship Attribution

INSTITUTE OF CZECH LITERATURE  
KAROLINUM PRESS

This PDF includes a chapter from the following book:  
Versification and Authorship Attribution  
© Petr Plecháč, 2021

## References

Petr Plecháč  
Institute of Czech Literature, Czech Academy of Sciences  
e-mail: plechac@ucl.cas.cz

This work is licensed under a Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.14712/9788024648903.6>

# References

- Abney, S. (2007). *Semisupervised Learning for Computational Linguistics*. Boca Raton, London and New York: CRC.
- Al-Falahi, A., Ramdani, M. and Bellafkih, M. (2017). Machine learning for authorship attribution in Arabic poetry. *International Journal of Future Computer and Communication*, 6(2), 42–46. doi: 10.18178/ijfcc.2017.6.2.486.
- Argamon, S. (2008). Interpreting Burrows's Delta: Geometric and probabilistic foundations. *Literary and Linguistic Computing*, 23(2), 131–147. doi: 10.1093/llc/fqn003.
- Baayen, R. H. (2001). *Word Frequency Distributions*. Dordrecht: Kluwer Academic Publishers. doi: 10.1007/978-94-010-0844-0.
- Bobenhausen, K. (2011). The Metricalizer. Automated metrical markup for German poetry. In C. Küper (ed.), *Current Trends in Metrical Analysis*. Frankfurt am Main et al.: Peter Lang, 119–131. doi: 10.3726/978-3-653-00980-4.
- Bobenhausen, K. and Hammerich, B. (2015). Métrique littéraire, métrique linguistique et métrique algorithmique de l'allemand mises en jeu dans le programme Metricalizer<sup>2</sup>. *Langages*, 199, 67–87. doi: 10.3917/lang.199.0067.
- Boyle, R. (1882). Massinger and *The Two Noble Kinsmen*. *New Shakspeare Society Transactions 1880–1885*, 371–399.
- Burrows, J. F. (1989). “An ocean where each kind”: Statistical analysis and some major determinants of literary style. *Computers and the Humanities*, 23(4–5), 309–321. doi: 10.1007/BF02176636.
- Burrows, J. F. (2002) “Delta”: A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3), 267–287. doi: 10.1093/llc/17.3.267.
- Burrows, J. F. (2003). Questions of authorship: Attribution and beyond. *Computers and the Humanities*, 37(1), 5–32. doi: 10.1023/A:1021814530952.
- Burrows, J. F. and Hassall, A. J. (1988). *Anna Boleyn and the Authenticity of Fielding's Feminine Narrative*. *Eighteenth-Century Studies*, 21(4), 427–453. doi: 10.2307/2738901.
- Craig, H. and Kinney, A. F. (2009). *Shakespeare, Computers and the Mystery of Authorship*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511605437.
- Čech, R. and Popescu, I.-I. and Altmann, G. (2011). Euphony in Slovak lyric poetry. *Glottometrics*, 22, 5–16.
- Červenka, M. (1998). Máchovo místo ve vývoji českého verše. *Česká literatura*, 46(5), 485–489.
- Červenka, M. and Sgallová, K. (1978). Český verš. In Z. Kopczyńska and L. Pszczołowska (eds.), *Słowińska metryka porównawcza I. Słownik rytmiczny i sposoby jego wykorzystania*. Wrocław et al.: Ossolineum, 45–94.

- De Morgan, A. (1851/1882). To Rev. W. Heald, Aug. 18, 1851 [letter]. In S. E. de Morgan (ed.), *Memoir of Augustus de Morgan*. London: Longmans, Green, & Co., 214–216.
- Diederich, J., Kindermann, J., Leopold, E. and Paass, G. (2003). Authorship attribution with support vector machines. *Applied Intelligence*, 19(1), 109–123. doi: 10.1023/A:1023824908771.
- Dobritsyn, A. (2016). Rhythmic entropy as a measure of rhythmic diversity (The example of Russian iambic tetrameter). *Studia Metrica et Poetica*, 3(1), 33–52. doi: 10.12697/smp.2016.3.1.02.
- Eddy, H. (1887). The characteristic curve of composition. *Science*, 9(216), 297. doi: 10.1126/science.ns-9.214S.237.
- Eder, M. (2011). Style markers in authorship attribution. A cross-language study of the authorial fingerprint. *Studies in Polish Linguistics*, 6, 99–114.
- Eder, M. (2016). Rolling stylometry. *Digital Scholarship in the Humanities*, 31(3), 457–469. doi: 10.1093/llc/fqv010.
- Eisen, M., Riberio, A., Segarra, S. and Egan, G. (2017). Stylometric analysis of early modern period English plays. *Digital Scholarship in the Humanities*, 33(3), 500–528. doi: 10.1093/llc/fqx059.
- Farnham, W. (1916). Colloquial contractions in Beaumont, Fletcher, Massinger and Shakespeare as a test of authorship. *Publications of the Modern Language Association of America*, 31, 326–358. doi: 10.1632/456960.
- Fleay, F. G. (1874a). On the authorship of *The Taming of the Shrew*. *Transactions of the New Shakspeare Society*, 1, 85–129.
- Fleay, F. G. (1874b). On the authorship of *Timon of Athens*. *Transactions of the New Shakspeare Society*, 1, 130–194.
- Fleay, F. G. (1874c). A fresh confirmation of Mr. Spedding's division and date of the play of *Henry VIII*. *Transactions of the New Shakspeare Society*, 1, appendix 23.
- Fleay, F. G. (1874d). Mr. Hickson's division of *The Two Noble Kinsmen*, confirmed by Metrical Tests. *Transactions of the New Shakspeare Society*, 1, appendix 61–64.
- Forstall, C. W., Jacobson, S. L. and Scheirer, W. J. (2011). Evidence of intertextuality: investigating Paul the Deacon's *Angustae Vitae*. *Literary and Linguistic Computing*, 26(3), 285–296. doi: 10.1093/llc/fqr029.
- Forstall, C. W. and Scheirer, W. J. (2010). A statistical stylistic study of Latin elegiac couplets. 2010 Chicago Colloquium on Digital Humanities and Computer Science Abstracts, Evanston: Northwestern University, 21–22. 11.
- Fucks, W. (1952). On mathematical analysis of style. *Biometrika*, 39(1–2): 122–129. doi: 10.1093/biomet/39.1-2.122.
- Furnivall, F. J. (1874a). The founder's prospectus of the New Shakspeare Society. *Transactions of the New Shakspeare Society*, 1, appendix 6–7.
- Furnivall, F. J. (1874b). Another fresh confirmation of Mr. Spedding's division and date of the play of *Henry VIII*. *Transactions of the New Shakspeare Society*, 1, appendix 24.
- Furnivall, F. J. (1874c). Mr. Hickson's division of *The Two Noble Kinsmen*, confirmed by the stop-line test. *Transactions of the New Shakspeare Society*, 1, appendix 64–65.
- Gasparov, M. and Tarlinskaja, M. (2008). The Linguistics of Verse. *The Slavic and East European Journal*, 52(2), 198–207.
- Giesbrecht, E. and Evert, S. (2009). Is Part-of-Speech Tagging a Solved Task? An Evaluation of POS Taggers for the German Web as Corpus. In I. Alegria, I. Leturia and S. Sharoff (eds.), *Proceedings of the 5th Web as Corpus Workshop (WAC5)*. San Sebastian.

- Göhring, A. (2009). Spanish Expansion of a Parallel Treebank. Master's thesis. Universität Zürich.
- Grieve, J. (2005). Quantitative authorship attribution: A history and an evaluation of techniques. Master's thesis. Simon Fraser University.
- Grzybek, P. (2014). The emergence of stylometry: prolegomena to the history of term and concept. In: Kroó, Katalin; Torop, Peeter (eds.), *Text within Text — Culture within Culture*. Budapest, Tartu: L'Harmattan, 58–75.
- Hajič, J. (2004). *Disambiguation of Rich Inflection (Computational Morphology of Czech)*. Prague: Karolinum.
- Halvani, O., Winter, C. and Graner, L. (2019). Assessing the Applicability of Authorship Verification Methods. In *Proceedings of the 14th International Conference on Availability, Reliability and Security (ARES '19)*. New York: Association for Computing Machinery, 1–10. doi: 10.1145/3339252.3340508.
- Hart, A. (1934). Shakespeare and the vocabulary of *The Two Noble Kinsmen*. *The Review of English Studies*, 10(39), 274–287.
- Herdan, G. (1956). Chaucer's authorship of the *Equatorie of the Planetis*: The use of romance vocabulary as evidence. *Language*, 32(2), 254–259.
- Hickson, S. (1847). The shares of Shakespeare and Fletcher in *The Two Noble Kinsmen*. *The Westminster and Foreign Quarterly Review*, 47, 59–48
- Holmes, D. I. (1998). The evolution of stylometry in humanities scholarship. *Literary and Linguistic Computing*, 13(3), 111–117. doi: 10.1093/lc/13.3.111.
- Hoover, D. L. (2003). Another perspective on vocabulary richness. *Computers and the Humanities*, 37(2), 151–178. doi: 10.1023/A:1022673822140.
- Hoover, D. L. (2004a). Testing Burrows's Delta. *Literary and Linguistic Computing*, 19(4), 453–475. doi: 10.1093/lc/19.4.453.
- Hoover, D. L. (2004b). Delta Prime? *Literary and Linguistic Computing*, 19(4), 477–495. doi: 10.1093/lc/19.4.477.
- Horsmann, T., Erbs, N. and Zesch, T. (2015). Fast or accurate? A comparative evaluation of PoS tagging models. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL-2015)*.
- Horton, T. B. (1987). *The Effectiveness of the Stylometry of Function Words in Discriminating between Shakespeare and Fletcher*. Dissertation. Edinburgh: University of Edinburgh.
- Hoy, C. (1962). The Shares of Fletcher and his collaborators in the Beaumont and Fletcher Canon VII. *Studies in Bibliography*, 15, 71–90.
- Iliushin, A. A. (1978). *Poeziia dekabrista Gavriila Stepanovitcha Batenkova*. Moscow: Izdatelstvo MGU.
- Iliushin, E. A. (ed.). (2020). *Izbrannyye stikhotvornye proizvedeniya*. Moscow: Common Place.
- Ingram, J. K. (1874). On the “weak endings” of Shakspeare, with some account of the history of the verse tests in general. *Transactions of the New Shakspeare Society*, 1, 442–456.
- Jakobson, R. (1938/1995). K popisu Máchova verše. In idem: *Poetická funkce*. Jinočany: H & H, 427–476,
- Jannidis, F., Pielström, S., Schöch, C. and Vitt, T. (2015). Improving Burrows' Delta. An empirical evaluation of text distance measures. *Digital Humanities Conference 2015*, Sydney.
- Jirát, V. (1931–1932). Hudebnost Máchova rýmu. *Časopis pro moderní filologii* 18(1–2), 24–34 & 147–157.

- Juola, P. (2006). Authorship attribution. *Foundations and Trends in Information Retrieval*, 1(3), 233–334. doi: 10.1561/1500000005.
- Kestemont, M. and Haverals, W. (2018). Metrical analyses of medieval Dutch poetry for the purpose of genre and authorship analysis [conference paper]. *Plotting Poetry II: Bringing Deep Learning to Computational Poetry Analysis*. Berlin: Freie Universität Berlin, 12.–14. 9. 2018.
- Koppel, M. and Schler, J. (2004). Authorship verification as a one-class classification problem. *Proceedings of the 21st International Conference on Machine Learning*. New York: ACM, 62–68. doi: 10.1145/1015330.1015448.
- Koppel, M., Schler, J. and Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the Association for Information Science and Technology*, 60(1), 9–26. doi: 10.1002/asi.20961.
- Koppel, M., Schler, J. and Bonchek-Dokow, E. (2007) Measuring Differentiability: Unmasking Pseudonymous Authors. *Journal of Machine Learning Research*, 8, 1261–1276.
- Larsen, W. A., Rencher, A. C. and Layton, T. (1980). Who wrote the *Book of Mormon*? An analysis of wordprints. *BYU Studies Quarterly*, 20(3), 225–251.
- Ledger, G. and Merriam, T. (1994). Shakespeare, Fletcher, and *The Two Noble Kinsmen*. *Literary and Linguistic Computing*, 9(3), 235–248. doi: 10.1093/lc/9.3.235.
- Lotman, J. M. and Lotman, M. (1986). Vokrug desjatoj glavy “Evgenija Onegina”. In N. N. Petrunina (ed.), *Pushkin: Issledovanija i materialy XII*. Moscow/Leningrad: Nauka, 124–151.
- Malone, Edmond. (1787/1803). A dissertation on parts one, two and three of *Henry the Sixth* tending to show that those plays were not written originally by Shakespeare. In *Plays of William Shakespeare 14*. London, 219–263.
- Mascol, C. (1888a). Curves of pauline and pseudo-pauline style I. *Unitarian Review*, 30, 452–460.
- Mascol, C. (1888b). Curves of Pauline and Pseudo-Pauline style II. *Unitarian Review*, 30, 539–546.
- Matthews, R. and Merriam, T. (1993). Neural computation in stylometry I; An application to the works of Shakespeare and Fletcher. *Literary and Linguistic Computing*, 8(4), 203–209. doi: 10.1093/lc/8.4.203.
- Mendenhall, T. C. (1887). The characteristic curves of composition. *Science*, 9(214), 237–249.
- Mendenhall, T. C. (1901). A mechanical solution to a literary problem. *Popular Science Monthly*, 9, 97–110.
- Mikros, G. K. and Perifanos, K. A. (2013). Authorship attribution in Greek tweets using author’s multilevel  $n$ -gram profile. In *Papers from the 2013 AAAI Spring Symposium*. “Analyzing Microtext”, 25–27 March 2013. Palo Alto: AAAI Press, 17–23.
- Mittmann, A., Pergher, P. H. and dos Santos, A. L. 2019. What rhythmic signature says about poetic corpora. In P. Plecháč, B. Scherr, T. Skulacheva, H. Bermúdez-Sabel and R. Kolár (eds.), *Quantitative Approaches to Versification*. Prague: ICL, 153–172.
- Moretti, F. (2013). *Distant Reading*. NY: Verso.
- Mosteller, F. and Wallace D. L. (1964). *Inference and Disputed Authorship*. Reading: Addison-Wesley.
- Nagy, B. (2021). Metre as a stylometric feature in Latin hexameter poetry. *Digital Scholarship in the Humanities [advance articles]*. doi: 10.1093/lc/fqaa043.
- Navarro-Colorado, B. (2015). A computational linguistic approach to Spanish Golden Age sonnets: metrical and semantic aspects. *Computational Linguistics for Literature NAACL 2015, Denver (Co)*. doi: 10.3115/v1/W15-0712.

- Navarro-Colorado, B. (2017). A metrical scansion system for fixed-metre Spanish poetry. *Digital Scholarship in the Humanities*, 33(1), 112–127. doi: 10.1093/llc/fqx009.
- Navarro-Colorado, B., Ribes-Lafoz, M. and Sánchez, N. (2016). Metrical annotation of a large corpus of Spanish sonnets. Representation, scansion and evaluation. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). Portorož.
- Oliphant, E. H. C. (1891). The Works of Beaumont and Fletcher. *Englische Studien* 15, 321–360.
- Oras, A. (1953). “Extra monosyllables” in *Henry VIII* and the problem of authorship. *Journal of English and Germanic Philology*, 52, 198–213.
- Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10(3), 61–74.
- Plecháč, P. (2016). Czech verse processing system KVĚTA: Phonetic and metrical components. *Glottotheory*, 7, 159–174. doi: 10.1515/glot-2016-0013.
- Plecháč, P. (2018). A collocation-driven method of discovering rhymes (in Czech, English, and French poetry). In M. Fidler and V. Cvrček (eds.), *Taming the Corpus. From Inflection and Lexis to Interpretation*. Cham: Springer, 79–95. doi: 10.1007/978-3-319-98017-1\_5.
- Plecháč, P. (2020). Relative contributions of Shakespeare and Fletcher in *Henry VIII*: An analysis based on most frequent words and most frequent rhythmic patterns. *Digital Scholarship in the Humanities* [advance articles]. doi: 10.1093/llc/fqaa032.
- Plecháč, P. and Birnbaum, D. (2019). Assessing the reliability of stress as a feature of authorship attribution in syllabic and accentual syllabic verse. In P. Plecháč, B. P. Scherr, T. Skulacheva, H. Bermúdez-Sabel, R. Kolár (eds.), *Quantitative Approaches to Versification*. Prague: ICL CAS, 201–210.
- Plecháč, P., Bobenhausen, K. and Hammerich, B. (2018). Versification and authorship attribution. Pilot study on Czech, German, Spanish, and English Poetry. *Studia Metrica et Poetica*, 5(2), 29–54. doi: 10.12697/smp.2018.5.2.02.
- Plecháč, P., Kolár, R. (2015). The Corpus of Czech Verse. *Studia Metrica et Poetica*, 2(1), 107–118. doi: 10.12697/smp.2015.2.1.05.
- Porter, J. D. (2018). Popularity/Prestige. Pamphlets of Stanford Literary Lab, 17.
- Rybicki, J. and Eder, M. (2011). Deeper Delta across genres and languages: Do we really need the most frequent words? *Literary and Linguistic Computing*, 26(3), 315–321. doi: 10.1093/llc/fqr031.
- Savoy, J. (2020). Advanced models for stylometric application. In: *Machine Learning Methods for Stylometry. Authorship Attribution and Author Profiling*. Cham: Springer, 153–187. doi: 10.1007/978-3-030-53360-1\_7.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*. Manchester.
- Segarra, S., Eisen, M. and Riberio, A. (2013). Authorship attribution using function words adjacency networks. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5563–5567. doi: 10.1109/ICASSP.2013.6638728.
- Shapir, M. I. (1997). Fenomen Batenkova i problema mistifikatsii (lingvistikhovedcheskij aspekt 1–2). *Philologica*, 4, 85–144.
- Shapir, M. I. (1998). Fenomen Batenkova i problema mistifikatsii (lingvistikhovedcheskij aspekt 3–4). *Philologica*, 5, 49–132.
- Shapir, M. I. (2000). Fenomen Batenkova i problema mistifikatsii. In M. I. Shapir. *Universum versus: Iazyk — stikh — smysl v russkoi poezii XVIII-XIX vekov*, 1. Moscow: Yazyki russkoi kultury, 335–443.

- Sherman, L. A. (1888). Some observations upon sentence-length in English prose. *The University of Nebraska Studies*, 1(4), 337–366.
- Skoumalová, H. (2011). Porovnání úspěšnosti tagování korpusu. In V. Petkevič and A. Rosen (eds.), *Korpusová lingvistika Praha 2011/3. Gramatika a značkování korpusů*. Prague: Nakladatelství Lidové Noviny, 199–207.
- Smith, P. W. H. and Aldridge, W. (2011). Improving authorship attribution: Optimizing Burrows' Delta method. *Journal of Quantitative Linguistics*, 18(1), 63–88. doi: 10.1080/09296174.2011.533591.
- Sobchuk O. (2018). *Charting Artistic Evolution: An Essay in Theory*. Dissertation. Tartu: University of Tartu Press, 2018.
- Spalding, W. (1833). *A Letter on Shakespeare's Authorship of The Two Noble Kinsmen*. Edinburgh: A. & C. Black.
- Spedding, J. (1850). Who wrote Shakespeare's *Henry VIII*? *The Gentlemen's Magazine*, 115–123.
- Spoustová, D., Hajič, J., Votrubec, J., Krbec, P. and Květoň, P. (2007). The best of two worlds: Cooperation of statistical and rule-based taggers for Czech. *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing. ACL*, 67–74.
- Tabata, T. (2012). Approaching Dickens' style through random forests. *Digital Humanities 2012: Conference Abstracts*. Hamburg: Universität Hamburg, 388–391.
- Tarlinkskaja, M. (1987). *Shakespeare's Verse: Iambic Pentameter and the Poet's Idiosyncrasies*. New York: Peter Lang.
- Tarlinkskaja, M. (2014). *Shakespeare and the Versification of English Drama, 1561–1642*. Farnham et al.: Ashgate.
- Thorndike, A. H. (1901). *The Influence of Beaumont and Fletcher on Shakespeare*. Worcester: Oliver B. Wood.
- Tomashevsky, B. V. (1923/2008). Pjatistopnyj jamb Pushkina. In *Izbrannye raboty o stikke*. Moscow & Sankt Peterburg: Akademija, 140–242.
- Vickers, B. (2004). *Shakespeare, Co-Author. A Historical Study of Five Collaborative Plays*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199269167.001.0001.
- Weber, H. (1812). Observations on the participation of Shakespeare in *The Two Noble Kinsmen*. In H. Weber (ed.), *The Works of Beaumont and Fletcher in Fourteen Volumes* 13. Edinburgh: J. Ballantyne & Co., 151–169.
- Williams, C. B. (1975). Mendenhall's studies of word-length distribution in the works of Shakespeare and Bacon. *Biometrika*, 62(1), 207–212. doi: 10.1093/biomet/62.1.207.
- Yule, G. U. (1939). On sentence-length as a statistical characteristic of style in prose, with application to two cases of disputed authorship. *Biometrika*, 30, 363–390.
- Yule, G. U. (1944). *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.
- Zhao, Y. and Zobel, J. (2005). Effective and scalable authorship attribution using function words. In G. G. Lee et al. (eds.), *Information Retrieval Technology. AIRS 2005. Lecture Notes in Computer Science*. Berlin and Heidelberg: Springer, 174–189. doi: 10.1007/11562382\_14.
- Zipf, G. K. (1932). *Selected Studies on the Principle of Relative Frequency in Language*. Cambridge, MA: Harvard University Press.