**ATIVIDADES ACADÊMICAS – 2024/2**

**Area de concentração:** Linguística Aplicada
**Disciplina: Stylometry: Computational stylistics**
**Código:** LIG 945 – B **Carga Horária (1 crédito = 15 h/a):** 15 horas
**Professor(es):** Adriana Pagano; Maciej Eder
**Modalidade:** ☒ Presencial ☐ Semipresencial ☐ Online **Vagas:** 20
**Período da disciplina (para disciplinas de 15h, 30h e 45h):** 21/10 a 25/10
**Dia da semana:** segunda a sexta **Horário:** 14 a 17:40

---

**Ementa:**
Stylometry, or the study of measurable features of (literary) style, such as sentence length, vocabulary richness and various frequencies (of words, word lengths, word forms, etc.), has been around at least since the middle of the 19th century, and has found numerous applications in authorship attribution research. These applications are based on the belief that there exist such conscious or unconscious elements of personal style that can help detect the true author of an anonymous text. But even more interesting research questions arise beyond bare authorship attribution: patterns of stylometric similarity and difference also provide new insights into relationships between different books by the same author; between books by different authors; between authors differing in terms of chronology or gender; between translations of the same author or group of authors; helping, in turn, to find new ways of looking at works that seem to have been studied from all possible perspectives.

**Programa:**

This will be achieved by placing at the participants' disposal some of the more useful stylometric tools and methods, from simple wordlist-making to custom-made multivariate analyses of word and phrase frequencies, for use in their individual projects. The tools are based in the R statistical programming environment, but they have been given user-friendly interfaces, so no expert knowledge of R in particular, or of programming in general is required. The texts used for the workshops will be provided by the instructors and the participants are encouraged to bring their own; if necessary, the participants' individual corpora will be expanded as needed and as available (online or elsewhere). The texts will be literary, multilingual, and include both originals and translations.

**Bibliografia:**
BERGSMA, POST, YAROWSKY. Stylometric analysis of scientific articles. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Montréal, Canada. Association for Computational Linguistics, p. 327–337, 2012. 8
BISCHOFF, DECKERS, SCHLIEBS, THIES, HAGEN, STAMATATOS, STEIN, POTTHAST. The importance of suppressing domain style in authorship analysis. arXiv preprint 2005.14714., 2020.
CINKOVÁ, S; RYBICKI, J. Stylometry in a bilingual setup. In: *Proceedings of the 12th Language Resources and Evaluation Conference*. 977-984, 2020.

CRYSTAL, D. “Think on my words” : exploring Shakespeare’s language. Cambridge ; New York: Cambridge University Press, 2008.
EDER, M. Visualization in stylometry: Cluster analysis using networks. *Digital Scholarship in the Humanities*, v. 32, n. 1, p. 50–64, 2 dez. 2015.
EDER, M.; RYBICKI, J.; KESTEMONT, M. Stylometry with R: a package for computational text analysis. *The R Journal*, v. 8, n. 1, p. 107, 2016.
EISENSTEIN, SMITH, XING. Discovering sociolinguistic associations with structured sparsity. In *Proc. ACL*, p.1365–1374, 2011.
FALTÝNEK D, MATLACH V. Hapax remains: Regularity of low-frequency words in authorial texts. *Digital Scholarship in the Humanities*. v. 37, n.3, p.693-715, 2022.
GORMAN, R. Author identification of short texts using dependency treebanks without vocabulary, *Digital Scholarship in the Humanities*, v. 35, n. 4, , p. 812–825, December 2020.
HOŁOBUT, A; WOŹNIAK, M.; RYBICKI, J. Old questions, new answers: computational stylistics in audiovisual translation research. *Audiovisual Translation: Research and Use*, 203-216, 2017.
KOPPEL; ARGAMON; SHIMONI. 2003. Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, 17(4):401–412, 2003.
KOPPEL; SCHLER; ZIGDON. Determining an author’s native language by mining a text for errors. In *Proc. KDD*, p. 624–628, 2005.
LE; LANCASHIRE; HIRST; JOKEL. 2011. Longitudinal detection of dementia through lexical and syntactic changes in writing: A case study of three British novelists. *Literary and Linguistic Computing*, 26(4):435–461, 2011.
LEE, CH. How do machine translators measure up to human literary translators in stylometric tests? *Digital Scholarship in the Humanities*, v. 37, n. 3, p. 813–829, September 2022.
MOSTELLER; WALLACE Applied Bayesian and classical inference: the case of the federalist papers. Springer-Verlag, 1984.
OTT, CHOI, CARDIE, HAN-COCK. Finding deceptive opinion spam by any stretch of the imagination. In *Proc. ACL*, pages 309–319, 2011.
PAGANO A, FIGUEREDO G, LUKIN A. Measuring proximity between source and target texts: an exploratory study. In: Tuzzi A, Benesová M, Macutek J (ed.) *Recent Contributions to Quantitative Linguistics*. Berlin, München, Boston: De Gruyter Mouton; 2015. p.103-114, 2015.
PAGANO A, FIGUEREDO G, LUKIN A. Modelling proximity in a corpus of literary retranslations: A methodological proposal for clustering texts based on systemic-functional annotation of lexicogrammatical features. In Meng Ji (ed). *Empirical Translation Studies - Interdisciplinary Methodologies Explored.* London: Equinox, 2014.
PIASECKI, M.; WALKOWIAK, T.; EDER, M. Open Stylometric System WebSty: Integrated Language Processing, Analysis and Visualisation. *Computational Methods in Science and Technology*, v. 24, n. 1, p. 43–58, 31 mar. 2018.
RAO, PAUL, FINK, YAROWSKY, OATES, COPPERSMITH. Hierarchical Bayesian models for latent attribute detection in social media. In *Proc. ICWSM*, pages 598–601, 2011.
RYBICKI, J. Burrowing into Translation: Character Idiolects in Henryk Sienkiewicz’s Trilogy and its Two English Translations. *Digital Scholarship in the Humanities*, v. 21, n. 1, p. 91–103, 24 mar. 2005.
RYBICKI, J. Vive la différence: Tracing the (authorial) gender signal by multivariate analysis of word frequencies. *Digital Scholarship in the Humanities*, v. 31, n. 4, p. 746–761, 8 jul. 2015.
RYBICKI, J.; HEYDEL, M. The stylistics and stylometry of collaborative translation: Woolf’s Night and Day in Polish. *Literary and Linguistic Computing*, v. 28, n. 4, p. 708–717, 27 maio 2013.

RYBICKI, Jan. The great mystery of the (almost) invisible translator. Stylometry in translation. In: Oakes MP, Ji M (ed.) *Quantitative Methods in Corpus-based Translation Studies: A Practical Guide to Descriptive Translation Research*, 2012, John Benjamins Publishing; 2012. p. 231–248.
SAVOY, J. *Machine learning methods for stylometry: authorship attribution and author profiling*. Cham, Switzerland: Springer, 2020.

**Pré-requisitos:**
Proficiência intermediária em língua inglesa - o curso será ministrado em língua inglesa
Habilidade de uso de softwares de edição de texto e planilha

**Outras exigências:**

Dedicação plena ao curso nos cinco dias de duração do mesmo