J7   Anonymous Letters Examination: Where Forensic Linguistics Meets Handwriting Examination

Andrea Ledic, MS*, Forensic Science Centre "I. Vucetic," Zagreb 10000, CROATIA; Lidija Tepes Golubic, PhD, University of Applied Sciences Zagreb, Zagreb 10000, CROATIA; Petra Bago, PhD, Faculty of Humanities and Social Sciences, Zagreb, CROATIA

Learning Overview: After attending this presentation, attendees will have a better understanding of the contents and/or the way threatening letters are formulated when sent to the ruling party or an individual in power by their opponent.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by introducing a unique study in which written letters were analyzed regarding the scope of the threatening letter, the contents, sentence complexity and repetition of certain words, and formulations.

The aim of this paper is to examine the possibilities of profiling the author of the threatening letters by analyzing the utilized linguistic formulations and to cluster groups of anonymous texts under some potential common identity.

The assumption is that there is mostly one group of people who send politically driven threatening letters. Special attention should be given to the letters that are in some way "signed," which can then refer to the fact that the author is the same, but this does not necessarily have to be true. The importance of this research is that no one has dealt with this topic seriously in Croatia yet.

In the first phase, testing and analysis of the handwriting of a certain number of anonymous letters have been conducted in order to determine whether or not there is evidence that two pieces of handwriting have a common authorship (i.e., is there any evidence that two pieces of handwriting are written by the same person or not). Preliminary analysis was performed to check whether the collected anonymous letters met the professional standards of handwriting analysis regarding their quality and quantity, as is the standard practice under International Organization for Standardization (ISO) 17025. Next, a detailed examination of general and individual handwriting characteristics was performed in all anonymous letters individually. A stereomicroscope and video spectral comparator were used to compare handwriting samples, as well as capture images and measurements.[1]

Even though the scope of this research does not necessarily include the graphic analysis spectrum, the described preliminary analysis was necessary to perform the following forensic linguistic comparison of the achieved results. For the purpose of further research, the handwritten letters have been transferred into a digital format, partially for the purpose of performing a computer text analysis.[2-4] The initial result achieved by identifying the author based on the handwriting characteristics has been verified using the method of forensic linguistics.

After it was established that the author of all analyzed letters was the same, semi-automatic linguistic research of the letters was conducted in order to compare the results.[5,6] Based on the vocabulary used, word order in sentences, sentence complexity, used nouns, adjectives, and verbs, the intention was to determine whether the common text author could be established through such partially automatic linguistic analysis.[7]

Computational analysis showed that an average number of sentences in the letters was 13.538, with the shortest letter containing 9 sentences and the longest 25. When examining all the letters, an average number of tokens per sentence was 11.392. One letter contained an average of only 4.364 tokens per sentence, while the letter with the longest sentences contained an average of 19.280 tokens. The data was compared to two general corpora of Croatian language: the Croatian web corpus hrWac and the Croatian Language Corpus (CLC). hrWac (version 2.2) contains 1.3 billion tokens compiled from documents collected on the *hr* domain. In this corpus, an average sentence contains 17.952 tokens.[8,9] CLC is a corpus containing documents written in standard language and contains 100 million tokens. An average number of tokens per sentence in this corpus is 18.074. The computational analysis has shown that the subject sentences are shorter than the ones in hrWac and CLC by approximately six to seven words.

Finally, it should be emphasized that the present research is a case study. In order to draw a universal conclusion, the research should be conducted on a larger number of letters. An interesting aspect of the conducted research is the analysis of handwritten letters and the comparison with the ones obtained through a linguistic and computer-helped analysis of individual words, sentences, or a text.

Reference(s):
1. Ellen D. Scientific Examination of Documents—Method and Techniques, 3rd edition. Taylor & Francis. 2006.
2. Chaski, C.E. Who's At The Keyboard? Authorship Attribution in Digital Evidence Investigations. International Journal of Digital Evidence 4, 2005, no. 1: 1-13.
3. Chaski, C.E. Best Practices and Admissibility in Forensic Author Identification. Journal of Law & Policy, Brooklyn Law School, Brooklyn, New York, 2013.
4. Chaski, C.E. Author Identification in the Forensic Setting. In L. Solan and P. Tiersma (Eds.), The Oxford Handbook of Forensic Linguistics (pp. 333-372). Oxford: Oxford University Press, 2012.
5. Carpenter, R. The statistical profile of language behavior with Machiavellian intent or while experiencing caution and avoiding self-incrimination. In R. Rieber & W. Stewart (Eds.), The Language scientist as expert in the legal setting: Issues in forensic linguistics 1990, (pp. 5-18). New York: The New York Academy of Sciences.
6. Mršić G., Galeković J., Ledić A., Risović A., Škavić N. Forenzika dokumenata, novca i rukopisa, Hrvatska sveučilišna naklada, Zagreb 2014, ISBN 9789531692861.
7. Tepeš Golubić, L. Značenje kao jedan od aspekata forenzične lingvistike. Semantika prirodnog jezika i metajezik semantike. Hrvatsko društvo za primijenjenu lingvistiku, Zagreb – Split 2005. str. 755-759. ISBN 953-96391-6-6.
8. Brozović Rončević, D. et al. 2018. Croatian language corpus Riznica 0.1, Slovenian language resource repository CLARIN.SI, http://hdl.handle.net/11356/1180.
9. Ljubešić, N., Klubička, F. {bs, hr, sr} wac-web corpora of Bosnian, Croatian and Serbian. In Proceedings of the 9th web as corpus workshop (WaC-9), 2014, (pp. 29-35).

Forensic Linguistics, Handwriting Examination, Computational Analysis

*Presenting Author