

## A STATISTICAL ANOMALY IN THE FIRST BI-LINES OF SHAKESPEARE SONNETS

*M.B. Malyutov and D.V. Khmelev*

University of Toronto, North-Eastern University

A controversy on who wrote Shakespeare works dates back to the 18th century with its bibliography on 1500 pages by 1947. The documentary and literary reasons for rejecting the authorship of the Stratford descendant are summarized e.g. in “The great cryptogram” by I. Donnelly, 1888, 1969 AMS Press (see also I.Gililov, “Igra ob Uil’yame Shekspire”, these hypotheses are called *non-stratfordian*). In 20th century the famous poet and playwright Christopher Marlowe emerged as one of main candidates. His blank iambic pentameter remained the principal style of English verse for several centuries. While expecting his death sentence for heresy and let out on bail, he allegedly perished under very suspicious circumstances at the very time when the first poem of yet completely unknown Shakespeare appeared. The reasons to believe that his death was faked including the contents of Shakespeare sonnets which point out to their Marlowe’s authorship, and stylometric analysis showing that Marlowe’s and Shakespeare styles are either identical or the latter is a natural development of the first, are on the informative site <http://www2.prestel.co.uk/rej/>. Several quantitative stylometry studies of Sheakespeare-Marlow controversy were undertaken by English scientists T. Merriam and R. Matthews, confirming e.g. that “Richard II” has much in common with Marlow writings. In November 2002 a Florida linguist R. Ballantine sent us her decipherment of anagrams in works of Shakespeare. It is almost unbelievable that their author could pursue additional goal while writing the top class poetry. However, first announcing discoveries by anagrams was very popular those times (Galilei, Huigens, Kepler, Newton among many prominent authors). Hence we attempted testing presence of anagrams statistically. An outline of some of our results based on the computations of D. Khmelev, follows: 72% or 111 of 154 Shakespeare sonnets contain the case-insensitive set of letters ‘m’, ‘a’, ‘r’, ‘l’, ‘o’, ‘w’, ‘e’ among their first bi-lines. Only 58% of other bi-lines in the sonnets contain the same set.

The normal approximation to the binomial probability of this or larger deviation is the right standard normal tail at around 6.5 which is a practically impossible event. Perhaps this anomaly signals that the first bi-lines were specially designed to include this set of letters.

---

Department of Mathematics, North-Eastern University, Boston, U.S.A.  
E-mail: mltv@neu.edu

Department of Mathematics, University of Toronto, Toronto, Ontario, Canada, M5S 3G3  
E-mail: dkhmelev@math.toronto.edu