А. В. Н е к л ю д о в (Москва, МГТУ). **О статистических характеристиках литературных текстов.**

Важной задачей математической лингвистики является изучение статистических характеристик текстов [1]–[3]. В [4] был предложен метод статистического анализа богатства авторского словаря для определения авторства текстов. Богатство словаря в [4] характеризуется выборкой значений числа различных (попарно несовпадающих) слов (словоформ), приходящихся на каждый из выделяемых в тексте блоков по 500 слов. Для подтверждения (или опровержения) авторства сравниваются выборка, взятая из текстов возможного автора, и выборка из атрибутируемого текста. При помощи t-критерия Стьюдента проверяется гипотеза о близости средних значений двух выборок, что, согласно [4], позволяет принять (или опровергнуть) вывод о принадлежности обоих текстов одному автору. При этом в [4] остается неисследованным вопрос о применимости данного критерия к анализу литературных текстов, а именно — является ли среднее значение числа словоформ авторским инвариантом, или оно может резко меняться в разных произведениях одного автора (или даже внутри одного его произведения)? Представленный же в [4] статистический материал делает такой анализ возможным. Ниже кратко приведены его результаты (см. также [5]).

В [4] приведены статистические данные по текстам Ф. Д. Крюкова и М. А. Шолохова: два текста Ф. Д. Крюкова — сборник издания 1907 г. (Кр-1) и сборник 1914 г. (Кр-2); два «бесспорных» текста М. А. Шолохова — «Донские рассказы» (ДР), 1-я книга «Поднятой целины» (ПЦ-1); три «спорных» текста М. А. Шолохова — 1-я, 2-я и 4-я части «Тихого Дона» (соответственно ТД-1, ТД-2 и ТД-4). Для проверки корректности метода [4] сравним между собой тексты, заведомо принадлежащие одному автору — Ф. Д. Крюкову либо М. А. Шолохову, а также сравним между собой различные части «Тихого Дона».

Сравнение по методу [4] двух текстов Ф. Д. Крюкова (тексты 1 и 2) дает следующий парадоксальный результат: гипотеза о принадлежности текстов Кр-1 и Кр-2 одному автору должна быть отвергнута с достоверностью 0,9998, поскольку значение теста t_{12} больше критического уровня распределения Стьюдента с соответствующим (88) числом степеней свободы $T(0,9998;88)$ и данной достоверности: $t_{12} = 4,0 > T(0,9998;88) = 3,9$. Сравнение методом [4] «бесспорных» произведений М. А. Шолохова (тексты 3 и 4) между собой дает основание отвергнуть гипотезу о едином авторстве ДР и ПЦ-1 с достоверностью 0,98, т. к. $t_{34} = 2,44 > T(0,98;80) = 2,37$. Наконец, попарное сравнение 1-й, 2-й и 4-й частей (тексты 5, 6, 7) «Тихого Дона» по методу [4] позволяет отвергнуть гипотезу о едином авторстве в отношении ТД-1 и ТД-4 с достоверностью 0,9, т. к. $t_{57} = 1,674 > T(0,9;63) = 1,669$, а в отношении ТД-2 и ТД-4 — с достоверностью 0,91, т. к. $t_{67} = 1,74 > T(0,91;68) = 1,72$. И только для ТД-1 и ТД-2 выборки оказываются близкими. Таким образом, рассматриваемый в [4] метод анализа богатства словаря не является корректным, т. к. данный параметр не является устойчивым в рамках творчества одного автора (или даже одного литературного произведения) — в четырех случаях из пяти тексты заведомо одного автора (в двух случаях при этом — части одного романа) были атрибутированы как принадлежащие различным авторам с достоверностью от 0,9 до 0,9998. Следовательно, этот метод не может ни подтвердить, ни опровергнуть авторство литературных произведений.

В силу вышесказанного, приводимые далее и основанные на методе [4] некоторые достоверности отвергнуть единое авторство для соответствующих пар текстов носят исключительно справочный характер и не могут быть аргументом в споре об авторстве: ДР и ТД-1 — 0,97; ДР и ТД-2 — 0,97; ДР и ТД-4 — 0,9998; ТД-4 и ПЦ-1 — 0,91.

СПИСОК ЛИТЕРАТУРЫ

1. *Ахманова О. С.*, *Мельчук И. А.*, *Падучева Е. В.*, *Фрумкина Р. М.* О точных методах исследования языка. М.: Изд-во МГУ, 1961.
2. *Фрумкина Р. М.* Статистические методы изучения лексики. М.: Наука, 1964.
3. *Головин Б. Н.* Язык и статистика. М.: Просвещение, 1971.
4. *Хьетсо Г.*, *Густавссон С.*, *Бекман С.*, *Гил С.* Кто написал «Тихий Дон». М.: Книга, 1989.
5. http://tikhij-don.narod.ru/Analis-Words.htm.