

Обнаружение интерференции у русско-французских писателей-билингвов с помощью  
авторской атрибуции

Маргарита В. Макарова

*Лозаннский университет,*

*Лозанна, Швейцария, margarita.makarova@unil.ch*

Андрей Макин, Валерий Афанасьев, Владимир Федоровский, Люба Юргенсон и Егор Гран — французские писатели русского происхождения. Согласно гипотезе исследования, в их франкоязычных текстах 1990-х гг. можно обнаружить случаи интерференции, то есть изменения литературной речи на одном языке (французском) под воздействием элементов другого языка (русского) на лексическом, морфологическом, грамматическом и/или синтаксическом уровнях.

Цель исследования — выявить интерференцию «вручную», а затем прибегнуть к авторской атрибуции из области компьютерной стилометрии, чтобы подтвердить или опровергнуть изначальные предположения<sup>1</sup>. Корпус состоял из двух групп — «русско-французской» и «франко-французской»<sup>2</sup>. В первую группу вошли романы так называемых русско-французских авторов, а именно Андрея Макина, Валерия Афанасьева, Владимира Федоровского, Любы Юргенсон, Егора Грана<sup>3</sup>, а во вторую — произведения Марселя Пруста, Эмиля Золя, Оноре де Бальзака, Гюстава Флобера, франко-французских писателей<sup>4</sup>. Такой выбор корпуса объясняется, в частности, тем, что язык произведений русско-французских авторов является книжным и отсылает к языку текстов французских «классиков» XIX в. Чтобы сравнить произведения писателей-билингвов русско-французской группы не только с «классиками», но и с их современниками, и заполнить таким образом лакуны данного исследования, в

---

<sup>1</sup> Под стилометрией («Stylometry») понимается процесс классификации литературных текстов с помощью статистических и компьютерных методов с целью нахождения сходств между стилями различных писателей. Одним из основных разделов стилометрии является авторская атрибуция («Authorship Attribution»). Автор анонимного текста определяется благодаря набору данных — текстов, чьи авторы известны. Помимо академических исследований, определение авторства текста требуется для выявления плагиата, а также применяется в области криминалистики [Savoy 2020].

<sup>2</sup> Кавычки при написании данных прилагательных впоследствии будут опускаться для удобства чтения.

<sup>3</sup> *Confession d'un port-drapeau déchu, Au temps du fleuve Amour, Le Testament français, Le Crime d'Olga Arbélina, Requiem pour l'Est* [Makine 1992, 1994, 1995, 1998, 2000]; *Lettres sonores* [Afanassiev 1995]; *Les Deux sœurs* [Fedorovski 1997]; *Éducation nocturne* [Jurgenson 1994]; *Acné festival* [Gran 1999].

<sup>4</sup> *Le Côté de Guermantes* (1920), *Du côté de chez Swann* (1913), *Un Amour de Swann* (1913) М. Пруста, *Germinal* (1885) Э. Золя, *La Maison du chat-qui-pelote* (1829) О. де Бальзака, *Bouvard et Pécuchet* (1881) Г. Флобера. Книги были загружены с сайта проекта «Гутенберг».

дальнейшем планируется включить в корпус и тексты франко-французских авторов 1990-х гг. В качестве инструментов для анализа были применены методы из области статистики и машинного обучения. Эксперименты были реализованы на языке программирования Python.

Первым методом был статистический метод «Критерий хи-квадрат» [Kilgarriff 2001]. С помощью него были подтверждены две гипотезы: 1) лексический состав романов А. Макина не схож с лексическим составом романов М. Пруста, 2) лексический состав романов А. Макина близок к лексическому составу романов В. Афанасьева, В. Федоровского, Л. Юргенсон, Е. Грана, но отличается от лексического состава франко-французских художественных текстов М. Пруста, Э. Золя, О. де Бальзака, Г. Флобера.

Затем были использованы три метода машинного обучения (метод «К-ближайших соседей» [Fix, Hodges 1989], метод «Гребневой регрессии» [Hoerl, Kennard 2000] и метод «Опорных векторов» [Cortes, Vapnik 1995]). Первый из них служил для бинарной классификации, а два других — для нахождения лексем, представляющих собой случаи интерференции. Наконец, метод «Главных компонент» был применен для визуализации данных [Jolliffe 1986].

В результате классификации с помощью методов «Критерий хи-квадрат» и «К-ближайших соседей», а также визуализации данных, произведения вышеназванных писателей действительно образовали две группы, русско-французскую и франко-французскую. Однако это еще не доказывает наличие в них интерференции. Увидеть конкретные примеры, то есть отдельные лексемы, склонившие выбор программы в пользу того или иного автора при классификации (так называемые «слова с наибольшим весом»), удалось с помощью методов «Гребневой регрессии» и «Опорных векторов». Более того, было интересно рассмотреть контекст употребления таких слов в произведениях русско-французских авторов.

### *Литература*

Сайт проекта «Гутенберг» – Библиотека электронных книг. [Электронный ресурс] URL: <https://gutenberg.org> (01.09.2021).

Afanassiev 1995 – *Afanassiev V. Lettres sonores*. Paris: José Corti, 1995. 260 p.

Cortes, Vapnik 1995 – *Cortes C., Vapnik V. Support Vector Networks // Machine Learning*. 1995. Vol. 20. P. 273–297.

- Fédorovski 1997 – *Fédorovski V.* Les Deux sœurs ou l'art d'aimer. Paris: JC Lattès, 1997. 232 p.
- Fix, Hodges 1989 – *Fix E., Hodges J.L.* Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties // *International Statistical Review / Revue Internationale de Statistique*. 1989. Vol. 57, no. 3. P. 238–247.
- Gran 1999 – *Gran I.* Acné festival. Paris: P.O.L, 1999. 188 p.
- Hoerl, Kennard 2000 – *Hoerl A.E., Kennard R.W.* Ridge regression: Biased estimation for nonorthogonal problems // *Technometrics*. 2000. Vol. 42, no. 1. P. 80–86.
- Jolliffe 1986 – *Jolliffe I.* Principal Component Analysis. New York, Berlin: Springer Verlag, 1986. 271 p.
- Jurgenson 1994 – *Jurgenson L.* Éducation nocturne. Paris: Albin Michel, 1994. 390 p.
- Kilgarriff 2001 – *Kilgarriff A.* Comparing corpora // *International Journal of Corpus Linguistics*. 2001. Vol. 6, no. 1. P. 97–133.
- Makine 1992 – *Makine A.* Confession d'un port-drapeau déchu. Paris: Pierre Belfond, 1992. 160 p.
- Makine 1994 – *Makine A.* Au temps du fleuve Amour. Paris: Les édition du Félin, 1994. 209 p.
- Makine 1995 – *Makine A.* Le Testament français. Paris: Mercure de France, 1995. 342 p.
- Makine 1998 – *Makine A.* Le Crime d'Olga Arbélina. Paris: Mercure de France, 1998. 287 p.
- Makine 2000 – *Makine A.* Requiem pour l'Est. Paris: Mercure de France, 2000. 287 p.
- Savoy 2020 – *Savoy J.* Machine Learning Methods for Stylometry. Authorship Attribution and Author Profiling. Switzerland: Springer, 2020. p. 3–17.