

Фирсова Светлана Павловна, Сулова Анастасия Михайловна

## **АВТОМАТИЗИРОВАННЫЙ АНАЛИЗ РУКОПИСНОГО ТЕКСТА В ОБРАЗОВАТЕЛЬНОЙ ДЕЯТЕЛЬНОСТИ**

Статья рассматривает основные характеристики и области применения рукописного текста. При этом особое внимание уделяется образовательному контексту, а именно проведению конкурсов, направленных на повышение уровня коммуникативной компетентности и межкультурной коммуникации участников и связанных с написанием рукописных текстов. Представлены различные виды анализа рукописной информации. Предлагается новый эффективный метод распознавания рукописной информации на основе контурного анализа.

Адрес статьи: [www.gramota.net/materials/2/2018/3-1/51.html](http://www.gramota.net/materials/2/2018/3-1/51.html)

Источник

**Филологические науки. Вопросы теории и практики**

Тамбов: Грамота, 2018. № 3(81). Ч. 1. С. 188-190. ISSN 1997-2911.

Адрес журнала: [www.gramota.net/editions/2.html](http://www.gramota.net/editions/2.html)

Содержание данного номера журнала: [www.gramota.net/materials/2/2018/3-1/](http://www.gramota.net/materials/2/2018/3-1/)

**© Издательство "Грамота"**

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: [www.gramota.net](http://www.gramota.net)

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: [phil@gramota.net](mailto:phil@gramota.net)

**INTER-LINGUAL FRENCH-RUSSIAN HOMONYMS  
РЕЖИССЁР / RÉGISSEUR, ДУБЛЁР / DOUBLEUR, КАДР / CADRE**

**Tkacheva Anna Nikolaevna**, Ph. D. in Philology, Associate Professor  
*Saint-Petersburg State University of Film and Television*  
*Tkacheva-Ann@yandex.ru*

The article considers the phonetically similar French and Russian words *régisseur* / *режиссёр*, *doubleur* / *дублёр*, *cadre* / *кадр*. The Russian words were borrowed from French to nominate new inventions in the sphere of theatre and cinema. Semantics of the French words is more extensional due to continuous development of polysemy. The Russian and French words differ in lexical meaning and contexts of usage. The author argues that the words *режиссёр*, *дублёр*, *кадр* and *régisseur*, *doubleur*, *cadre* are inter-lingual homonyms due to discrepancy in semantics.

*Key words and phrases:* inter-lingual homonyms; acoustic similarity; semantic difference; borrowing; etymology; cultural contacts.

УДК 81'33

*Статья рассматривает основные характеристики и области применения рукописного текста. При этом особое внимание уделяется образовательному контексту, а именно проведению конкурсов, направленных на повышение уровня коммуникативной компетентности и межкультурной коммуникации участников и связанных с написанием рукописных текстов. Представлены различные виды анализа рукописной информации. Предлагается новый эффективный метод распознавания рукописной информации на основе контурного анализа.*

*Ключевые слова и фразы:* рукописный текст; коммуникативная компетенция; тотальный диктант; методы распознавания; контурный анализ.

**Фирсова Светлана Павловна**, к. пед. н., доцент

**Сулова Анастасия Михайловна**

*Поволжский государственный технологический университет, г. Йошкар-Ола*  
*svetlana\_firsova@inbox.ru; a\_m\_suslova@mail.ru*

**АВТОМАТИЗИРОВАННЫЙ АНАЛИЗ РУКОПИСНОГО ТЕКСТА  
В ОБРАЗОВАТЕЛЬНОЙ ДЕЯТЕЛЬНОСТИ**

В современной образовательной ситуации, характеризуемой формированием международного образовательного пространства, проблема повышения иноязычной коммуникативной компетенции и развития навыков межкультурной коммуникации и взаимодействия является актуальной для лингвистики и лингводидактики. В целях решения обозначенной проблемы организуются различные конкурсы и акции, связанные с написанием рукописных текстов. В качестве примера можно привести ежегодный тотальный диктант по русскому языку, который проводится для всех желающих и собирает порядка 200 тысяч человек более чем в 800 населенных пунктах в России и в 71 стране мира. Диктант пишется от руки лично каждым участником [4].

По аналогии с данным событием, по инициативе кафедры иностранных языков и лингвистики Поволжского государственного технологического университета с 2013 года успешно реализуется акция «Тотальный английский», включающая написание диктанта на английском языке на основе англоязычного аудиотекста в форме фрагмента известного произведения, объемом 1200-1500 знаков. Основной целью акции является популяризация английского языка в качестве естественного и необходимого языка общения XXI века, мотивация к его изучению и владению им на уровне мировых стандартов. При этом реализуются следующие образовательные задачи: развитие лингвистической компетенции студентов и школьников, повышение уровня коммуникативной, общей и информационной культуры участников акции; формирование навыков межкультурной толерантности и взаимодействия, содействие формированию активной жизненной позиции молодежи [6]. Практический результат образовательной акции заключается в расширении средств контроля и оценки в области лингвистического образования, а также повышении уровня мотивации и интереса к изучению иностранных языков за счет более эффективного подбора образовательных мероприятий, унифицированной оценки уровня лингвистической компетенции, масштаба участия обучаемых и экспертов. Акция проходит на нескольких образовательных площадках, ежегодно в ней принимают участие от 500 до 700 человек различного возраста и уровня образованности, обладающие различными индивидуальными особенностями написания английских букв от руки.

Необходимо отметить, что, несмотря на высокий уровень глобальной информатизации и автоматизации, рукописный текст представляет уникальный материал для анализа и исследования в ряде ситуаций. Так, помимо образовательной деятельности (конкурсы, акции, выполнение домашних заданий), текст, написанный от руки, становится предметом анализа в криминалистике, психологии, документоведении, т.е. в предметных областях, где необходима идентификация личности автора и автономности выполняемой работы. Таким образом, обращение к рукописному тексту представляет интерес для исследователей в области прикладной лингвистики.

Под рукописным текстом нами понимается текст, который выражает определенную мысль человека и фиксируется при помощи знаков, символов, написанных от руки. Рукописный текст обладает как общими, характерными для всех текстов, так и специфическими признаками. Так, в числе основных общих признаков, имеющих особое значение для анализа информации, написанной от руки, выделяются: информативность, членимость, когезия, завершенность [1]. К специальным признакам рукописного текста можно отнести особенности наклона и форму написания букв, топографические признаки (расположение текста и его составных частей на бумаге), мелкие привычки письма (выделение частей текста, исправления в тексте и способы их выполнения, нумерация страниц, глав, сокращение слов) [7].

Различные области применения рукописного текста актуализируют необходимость автоматизации процесса его анализа и последующей обработки в целях облегчения деятельности эксперта и снижения субъективного фактора при работе с большим массивом рукописной информации. Для облегчения работы эксперта, а также для уменьшения субъективной оценки и возможных ошибок разрабатываются средства и методы по решению задач автоматизации. Задача автоматизации анализа рукописного текста требует использования специальных методов и техник обработки визуальной информации, направленных на распознавание рукописного текста.

Следует подчеркнуть, что решением задачи распознавания рукописного текста исследователи в области лингвистики и искусственного интеллекта занимаются достаточно продолжительное время, но из-за трудностей работы с рукописями и индивидуальности почерка задача создания универсального программного комплекса, распознающего любой рукописный текст, остается нерешенной. Как было отмечено ранее, рукописный текст обладает определенным количеством общих и специальных признаков, которые зависят от области исследования. Поэтому для каждой области исследования разработано немало программ, решающих узконаправленные проблемы по анализу одного или двух признаков.

Так, например, алгоритм программы, разработанной А. В. Кучугановым и Г. В. Лапинской (Ижевский государственный технический университет) в 2006 году, заключается в следующем: изображение рукописи преобразовывают в векторное, в результате текст представляется в виде отрезков прямых и дуг. Исходные элементы изображения и эталонные образцы представляются в виде графа. Просматривая эталонные образцы и граф анализируемого текста, выполняя необходимые пометки при совпадении, осуществляется распознавание изображения рукописного текста. Данный алгоритм программы направлен на анализ только наклона букв в словах и требует длительной предобработки и расчета дополнительных параметров [2].

Следующий подход был реализован в рамках исследования А. А. Мозгового (Воронежский институт высоких технологий). На вход алгоритма подается очищенное от шумов и посторонних объектов изображение, содержащее исключительно текстовую информацию. При этом определяется плотность текста для вычисления предполагаемого расстояния между словами. Далее осуществляется поиск связанных друг с другом точек рекурсивным методом. Выполнив поиск связанных друг с другом точек, получают набор совокупностей точек с пересечением некоторых областей. В рассмотренном алгоритме для анализа используется расстояние между буквами и словами, также большое количество процедур обработки изображения и расчета дополнительных параметров [3].

Для автоматизированного анализа текста может быть использовано специально разработанное программное обеспечение, например "Transcribus". Данная программа успешно справляется с задачей распознавания текста, но требует длительного обучения для понимания стиля распознаваемых документов и тщательного редактирования текстовой информации, а также создания тематических библиотек для качественного распознавания [8].

Таким образом, основными недостатками существующих алгоритмов распознавания и анализа рукописного текста являются узконаправленность, необходимость проведения большого количества процедур обработки, отсутствие инструментов для верификации результатов. В связи с этим была поставлена исследовательская задача по разработке метода автоматизированного анализа текста.

Независимо от области применения текста, для его анализа в основном используют статистические методы и методы квантитивной лингвистики. Распознавание рукописного текста может быть осуществлено на основе узконаправленных методов анализа, из которых можно создать комплексное программное обеспечение, объединяющее описанные методы. Такое программное обеспечение может применяться для различных областей исследования и расширит возможности анализа, ограниченные направленностью методов.

Работа по разработке метода автоматизированного анализа рукописного текста включала: сбор необходимых для исследования данных; предварительную обработку анализируемого изображения; выделение контура изображения слова; создание базы эталонных контуров; проведение согласованной фильтрации, которая заключается в математическом сравнении двух контуров, эталонного и исследуемого; поиск диапазонов значений модуля согласованной фильтрации, приближенных к максимальному значению.

В основе предлагаемого нами метода лежит анализ наклона букв, размера букв и расстояния между ними [5].

Также метод инвариантен к углу поворота, то есть для данного метода не имеет значения, как отсканирован документ или под каким наклоном написан текст, распознавание будет осуществлено. Необходимо отметить, что метод является языкозависимым, то есть для распознавания документов на разных языках необходимо выбирать свою базу эталонных букв и слов.

В процессе реализации разработанного алгоритма на вход программы поступает отсканированное изображение рукописи (Рис. 1).

Затем происходит побуквенное сравнение входного изображения с эталонными образцами букв. Согласованная фильтрация заключается в сравнении двух контуров с помощью инструментов математического анализа. По максимальному значению на графике модуля согласованной фильтрации делается вывод о результате распознавания: если в сравниваемых объектах буквы совпадают, то модуль согласованной фильтрации покажет максимальное значение, что и будет означать успешное распознавание.

*claim oligarchs*

Рисунок 1. Исходные изображения слова

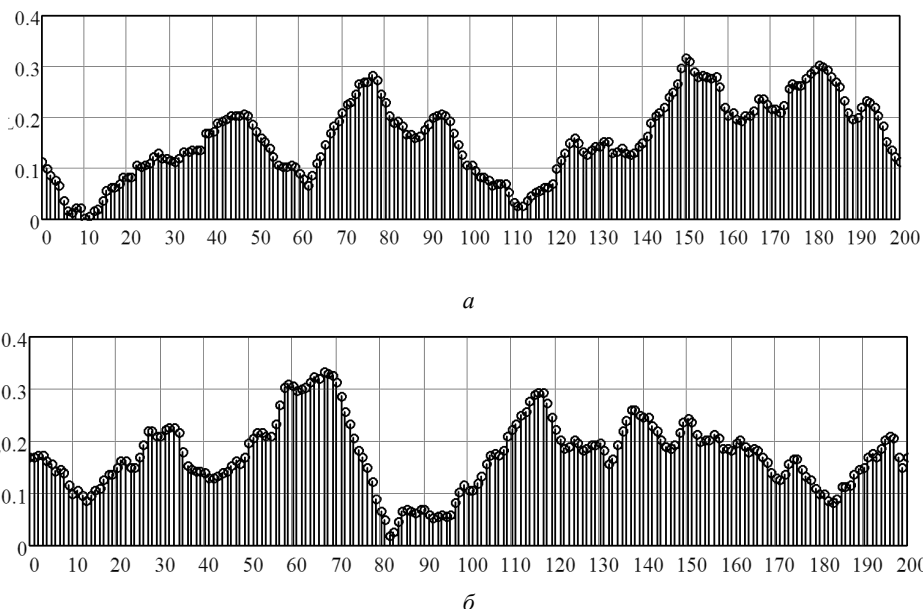


Рисунок 2. Графики модуля согласованной фильтрации для слов: а – claim; б – oligarchs

Основными преимуществами описанного алгоритма является исключение трудоемких и многоступенчатых вычислений и предварительного анализа текста.

Таким образом, разработанный алгоритм несет практическую значимость и может быть успешно применен в образовательной деятельности, позволяя осуществлять автоматизированное распознавание работ, написанных от руки, их анализ и проверку, что облегчит деятельность преподавателей, экспертов, исследователей. Метод, основанный на контурном анализе, может быть использован и как комплекс для осуществления распознавания текста, и как часть уже существующей программы для повышения ее эффективности.

#### Список источников

1. Гальперин И. Р. Текст как объект лингвистического исследования. Изд-е 4-е, стереотипное. М.: КомКнига, 2006. 144 с.
2. Кучуганов А. В., Лапинская Г. В. Распознавание рукописных текстов // Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам: мат-лы международной научной конференции / отв. ред. В. А. Баранов. Ижевск: ИжГТУ, 2006. С. 98-103.
3. Мозговой А. А. Проблемы извлечения рукописных слов из сканированных изображений // Моделирование, оптимизация и информационные технологии. 2013. № 1. С. 4-14.
4. Организаторы «Тотального диктанта» оценили количество участников в 2017 году [Электронный ресурс]. URL: <http://www.interfax.ru/culture/557458> (дата обращения: 20.10.2017).
5. Сулова А. М., Жимбоев М. М. Распознавание рукописного текста // Синергия наук. 2017. № 17. С. 67-71.
6. Фирсова С. П., Щеглова Н. Н. Тотальный диктант как средство повышения иноязычной коммуникативной компетенции студентов и школьников // Иностранные языки в школе. 2015. № 6. С. 59-62.
7. Чернов Ю. Г. Анализ почерка в работе с кадрами. СПб.: БХВ-Петербург, 2012. 310 с.
8. Transcribe. Collaborate. Share... [Электронный ресурс]. URL: <https://transkribus.eu/Transkribus/> (дата обращения: 24.10.2017).

#### AUTOMATED ANALYSIS OF A HANDWRITTEN TEXT IN EDUCATIONAL ACTIVITY

Firsova Svetlana Pavlovna, Ph. D. in Pedagogy, Associate Professor  
 Suslova Anastasiya Mikhailovna  
 Volga State University of Technology, Yoshkar-Ola  
 svetlana\_firsova@inbox.ru; a\_m\_suslova@mail.ru

The article examines the main characteristics and areas of handwritten text application. At the same time, special attention is paid to the educational context, namely the holding of contests aimed at increasing the level of communicative competence and intercultural communication of participants, and connected with writing texts. Various types of handwritten information analysis are presented. A new effective method for recognizing handwritten information based on a contour analysis is proposed.

*Key words and phrases:* handwritten text; communicative competence; total dictation; methods of recognition; contour analysis.