6. *Прищепа В. П.* Проблемы развития русской советской поэмы в 1960–1965 гг. (Некоторые тенденции развития жанров). Иркутск, 1985. 199 с.
7. *Рождественский Р. И.* Избранные произведения: В 2-х т. Стихотворения; Поэмы (1951–1966). М.: Худож. лит., 1979. Т. 1.
8. Рождественский Р. И. Избранные произведения: В 2-х т. Стихотворения; Поэмы (1965–1977); Песни. М.: Худож. лит., 1979. Т. 2. 472 с.

REFERENCES

1. *Gerasimenko A. P.* Problema haraktera v russkoj sovetskoj lirike 60-h godov: Dis. na soisk. uchen. step. kand. filol. nauk. M.: 1973. 172 s.
2. *Mal'gin A. V.* Besedy o poeme / Interv'ju s E. Evtushenko, R. Rozhdestvenskim, L. Ozerovym. M.: Obshch-vo "Znanie" RSFSR, 1990. 64 s.
3. *Mal'gin A. V.* Robert Rozhdestvenskij. Ocherk tvorchestva. M.: Hudozh. lit., 1990. 206 s.
4. *Mil' L.* Sluchaj rezonansa // Literaturnaja gazeta. 1968. 28 fevralja. S. 8.
5. *Mudrik M.* XX vek. Poety. Omsk. Omsk, 2008.
6. *Prishchepa V. P.* Problemy razvitija russkoj sovetskoj pojemy v 1960–1965 gg. (Nekotorye tendentsii razvitija zhanrov.) Irkutsk, 1985. 199 s.
7. *Rozhdestvenskij R. I.* Izbrannye proizvedenija: V 2-h t. T. 1. Stihotvorenija; Pojemy (1951–1966); Predisl. E. Sidorova. M.: Hudozh. lit., 1979.
8. *Rozhdestvenskij R. I.* Izbrannye proizvedenija: V 2-h t. T. 2. Stihotvorenija; Pojemy (1965–1977); Pesni. M.: Hudozh. lit., 1979, 472 s.

Е. И. Бойчук

ПРИМЕНЕНИЕ НАВЫКОВ КОМПЬЮТЕРНОЙ ОБРАБОТКИ РИТМА ТЕКСТА В ХОДЕ ПРАКТИЧЕСКИХ ЗАНЯТИЙ ПО ФРАНЦУЗСКОМУ ЯЗЫКУ

Рассматривается специфика применения компьютерных средств обработки текста с ориентированием на анализ ритмики французских художественных произведений. Авторами представлен алгоритм создаваемой компьютерной программы анализа ритма по основным языковым параметрам на фонетическом уровне ритмизации, а также описаны методологические способы их применения в ходе занятий по иностранному (французскому) языку. Основной функцией компьютерной программы «Анализ ритма французского текста» является вспомогательная функция, а именно выполнение механических операций по обработке таких ритмических средств, как аллитерация, ассонанс, выделение равных по количеству слогов ритмических единиц, рифмы.

Ключевые слова: компьютерная программа, обработка текста, анализ ритма, французский язык, метод.

E. Boichuk

APPLICATION OF COMPUTER TEXT-PROCESSING METHODS IN THE RHYTHM ANALYSIS WHILE TEACHING FRENCH

The object of this article is to consider the peculiarities of the application of computer means of text-processing implementing the rhythm analysis of French texts. The algorithm of rhythm analysis is fulfilled from the point of view of the main phonetic parameters. The most important

function of computer program "Rhythm analysis of French text" is the supporting and execution of mechanic operations with such means of rhythm as alliteration, assonance, rime and equal rhythm units. This article also represents the methods of program application while teaching French.

Keywords: computer program, text-processing, rhythm analysis, French, method.

Аспект исследования, описанный в данной статье, представлен одним из методов работы с текстом в рамках компьютерной лингвистики, в частности, таких ее направлений, как системная обработка, автоматическое извлечение фактов. Извлечение определенной информации из текста определяется конкретным коммуникативным намерением исследователя. Предлагаемая система обработки является проблемно-ориентированной, направленной на исследование ритмических характеристик французского текста.

В современной компьютерной лингвистике существует множество программ, направленных на исследование текста с точки зрения различных его аспектов. Так, существуют программы, позволяющие реализовать полный лингвистический анализ текстов на русском, английском, немецком и французском языках (ПОЛИГЛОТ, Russian Context Optimizer, Яndex. Server, Link Grammar Parser и др.). Модули лингвистического анализа включают в себя морфологию (словарную и бессловарную), синтаксис, семантику (включая толково-комбинаторные словари). Имеется модуль семантической рубрикации текстов. Однако зачастую анализ текста на таких уровнях лингвистического анализа, как графематический, морфологический, семантический — ограничивается использованием тематических синонимов, а также поиска по ключевым словам и формирования отчетов по частоте встречаемости слов в документах.

Наибольший интерес представляют программы, позволяющие дифференцировать идиостиль писателя, выявлять принадлежность текста тому или иному автору, а также программы, определяющие функциональные стили речи. В разряде таких программ «Лингвоанализатор» Д. В. Хмелева, «Худломер» Л. Делицына. Целью первой является определение близости любого из предлагаемых пользователем текстов к одному из авторских эталонов, определенных заранее и взятых из ресурсов Русской Фантастики. Программа анализирует входной текст и выдает имена трех писателей, которые могли бы быть его наиболее вероятными авторами. Кроме этого, программа находит три произведения каждого из авторов, которые наиболее близки данному тексту. Вторая программа производит автоматическую классификацию функциональных стилей текста (разговорного, художественного, газетно-информационного, научно-делового) на основе спектров длин слов.

Большой интерес представляет лингвистическая программа «Rhymes» Н. Кецариса, выполняющая поиск рифм на основе словаря А. А. Зализняка. При поиске программа реализует фонетическое сравнение слов с учетом ударения, позволяет находить для заданного слова синонимы и антонимы.

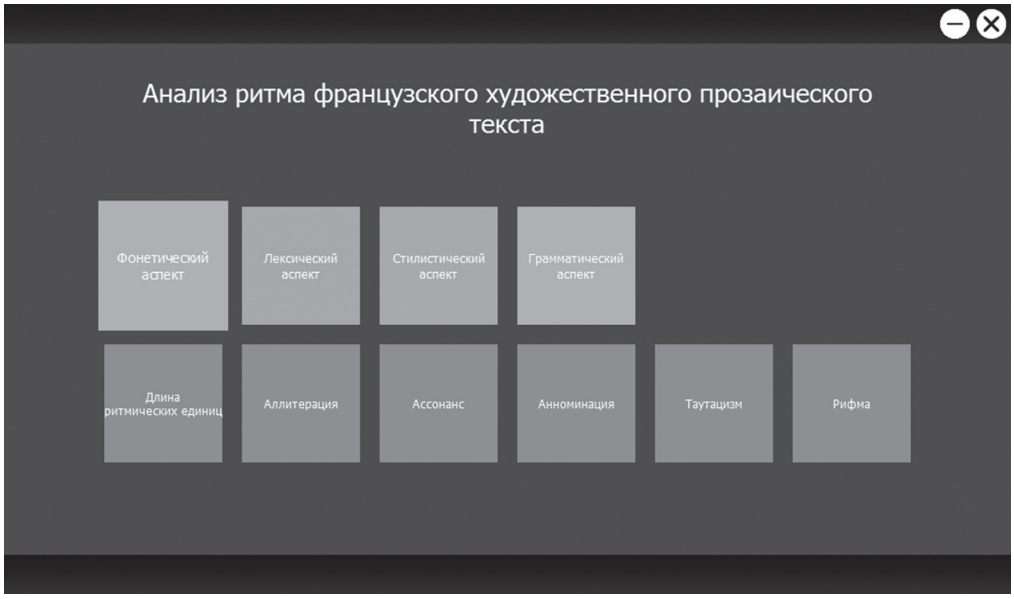
Алгоритм программы «Анализ ритма французского текста» разрабатывается автором данной статьи совместно с программистом Н. И. Кожемякиным. Программа позволяет комплексно подойти к проблеме проявления ритмических характеристик художественного текста. С практической точки зрения данные навыки работы с текстом применимы в ходе работы с аутентичными художественными как прозаическими, так и поэтическими текстами. Ритмический анализ, выполненный при помощи данной программы, позволяет глубже проникнуть в суть произведения, в идею автора, более полно представить рисуемые образы.

В рамках данной статьи описывается начальный этап работы над компьютерной лингвистической программой, позволяющей анализировать французские художественные прозаические тексты. Программа осуществляет парсинг введенного текста, т. е. разбор строки по определенным правилам, осуществляемый пошаговым методом. При парсинге исходный текст преобразуется в структуру данных, обычно — в дерево, которое отражает синтаксическую (в рамках информатики, не путать с синтаксисом в лингвистике) структуру входной последовательности и хорошо подходит для дальнейшей обработки [1].

Условия разбора текста описаны правилами произношения и слогоделения во французском языке. При создании программы использовались следующие языки программирования: QML (для написания интерфейса), JavaScript (для написания логики, связанной с интерфейсом (внешним видом программы)), HTML (для оформления текста), C++ и Qt (для создания функциональной части программы).

Анализ художественного текста осуществляется с точки зрения трех основных параметров: поиска фонетических, лексико-стилистических и грамматических средств ритмизации. Наиболее полно данный аспект представлен в монографии автора статьи [2].

Интерфейс программы выглядит следующим образом:



Как видно из скриншота, в рамках фонетического аспекта предлагается анализ таких средств реализации ритма, как длина ритмических единиц (выявление равенства или последовательности слогов), рифма, аллитерация, ассонанс, анноминация, таутацизм. Внутри тайла «рифма» существует деление на точные и неточные рифмы, позволяющее оценивать текст с точки зрения их характера в разных экранах ввода.

Лексико-стилистический аспект представлен следующими средствами: анафора, эпифора, симплока, а также различные типы повторов (простой, градационный, эпаналепсис, антанаклаза, мимезис, пситтацизм), частота употребления антонимов и синонимов.

Анализ грамматического аспекта ритмизации проводится на основе анализа морфемного состава слов, выявления однородных членов, характеризующихся наличием определенных графических символов, частотности предложений, общих по цели высказывания (вопрос, восклицание, умолчание (апозиопеза, апория)), особенности по-

рядка слов в предложении (инверсия, хиазм, симплока). Необходимо отметить, что программой выявляется частотность проявления перечисленных средств, что позволяет судить о периодичности их появления в тексте, а следовательно, о ритмичности текста. Наполнение прозаического текста теми или иными ритмическими средствами, в свою очередь, позволяет судить об особенностях языка того или иного писателя, его идиостиля.

Поиск и выявление перечисленных средств в тексте рассматриваются как типы его обработки. При этом анализируемый текст является динамическим, поскольку программа позволяет изменять существующие параметры при работе с ним. Необходимо отметить, что предлагаемая программа является вспомогательной при обработке художественного прозаического текста с точки зрения фонетического, лексико-стилистического и грамматического аспектов. Кроме того, ряд ритмических составляющих в рамках предложенных аспектов, зачастую зависящих от субъективных факторов, в частности от манеры чтения, от темпа речи, от стиля изложения, а также от возможностей компьютера, во внимание не принимаются. Среди таких средств, например, можно отметить интонационные характеристики звучащего текста (мелодика, темп, тембр, интенсивность, паузы, цезуры, сустентации (замедления)), синтаксический параллелизм, цепное нанизывание предложений и др.

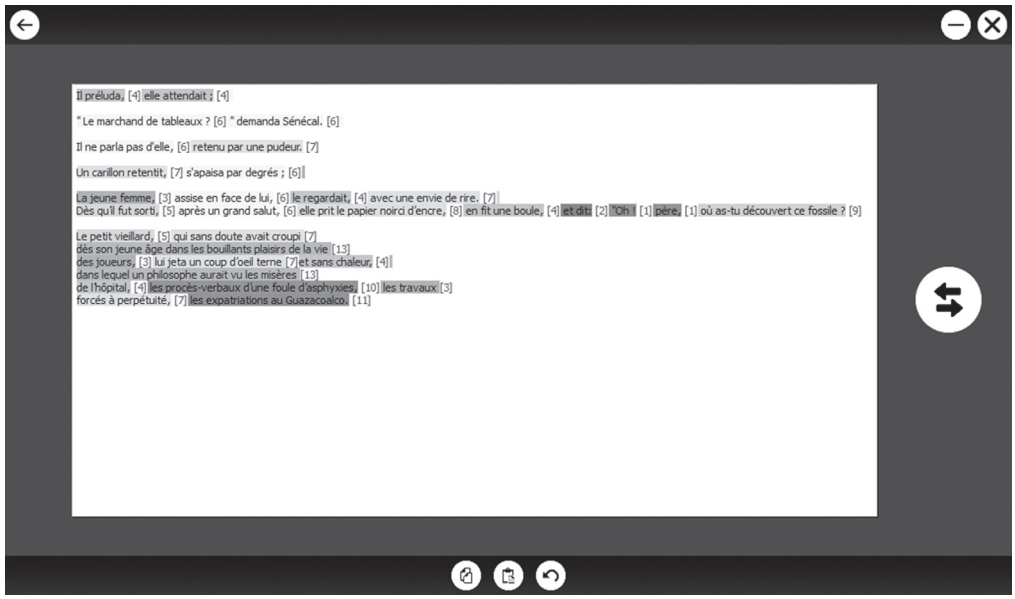
В рамках наиболее показательного с точки зрения ритма фонетического аспекта выявление большинства описанных выше средств сводится, в первую очередь, к делению на слоги предложенного программе текста. Правила, позволяющие программе делить текст на слоги, сформулировать довольно сложно, в первую очередь, в связи с тем, что компьютер воспринимает не звуки, фонемы или буквы, а знаки. При создании программы за основу принимались правила слогоделения, правила чтения, звуковые сочетания, дифтонги, которые должны распознаваться программой как один слог (iè, ieu, eai и др.), а также всевозможные позиции согласных, звучание которых зависит от ближайшего окружения.

В программе прописаны правила произношения **e** беглого, что составило большую трудность, поскольку четко регламентированных правил произношения этого звука нет, все зависит от стиля, от темпа и многих других факторов, в том числе и от индивидуальной манеры чтения. За основу были приняты правила произношения беглого **e** в позиции двух согласных перед третьей. Кроме того, программа прописывает случаи сокращений типа Mme, M., которые составляют по два и более слогов, в отличие от прописного варианта.

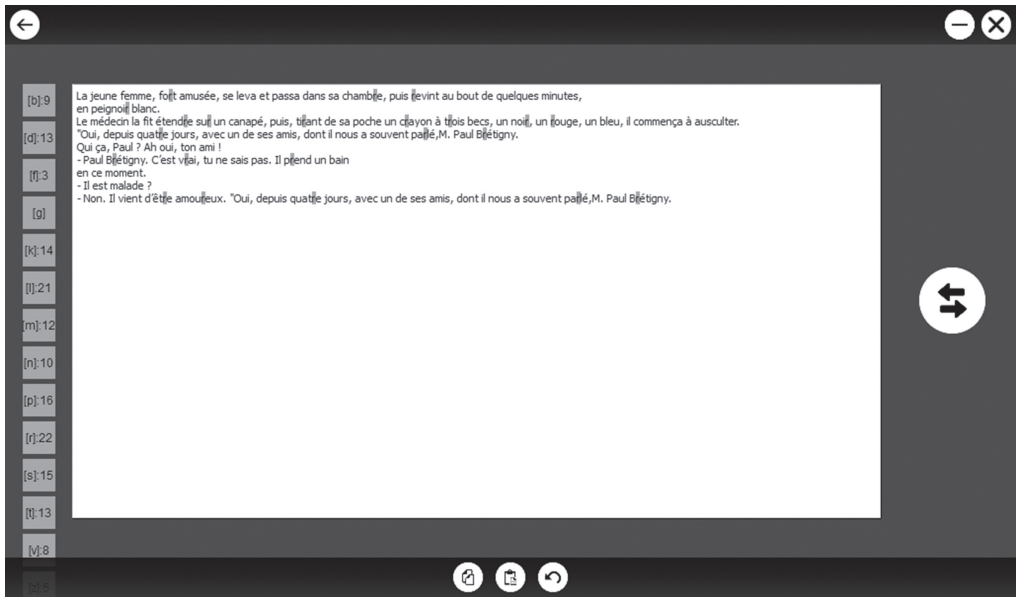
На основе этих и многих других введенных правил слогоделения прописывается алгоритм выделения ритмических единиц текста. В связи с тем, что выделение ритмических групп в рамках программы сложно подчинить общим правилам, существующим в фонетике, поскольку разграничение частей речи и членов предложения — задача для машины мало возможная, здесь были использованы правила деления на группы при помощи графических знаков, а именно запятых, двоеточий, точки с запятой, тире, а также союзов, как сочинительных, так и подчинительных. В рамках деления на ритмические единицы и нахождения равных по количеству слогов отрезков программа не различает конкретных звуков, а делит все существующие знаки на токены: пробелы, знаки препинания, союзы, а также знаки, соответствующие согласным и гласным буквам.

Безусловно, такое деление нельзя назвать точным, однако на данном этапе можно назвать это делением не на ритмические группы, а на некие ритмические единицы, облегчающие работу с ритмом текста. Деление на ритмические единицы и подсчет в них слогов позволяет выявить их равенство, прогрессии или последова-

тельность, что напрямую связано с ритмом текста: единицы, равные с точки зрения слогового состава или представляющие собой слоговую последовательность с разницей в один слог, являются наиболее ритмичными. Каждому количеству слогов в рамках единицы соответствует свой цвет в порядке увеличения или уменьшения интенсивности цвета в зависимости от количества слогов:



С целью выявления повторяющихся согласных и гласных звуков в определенном фрагменте текста были сформулированы правила, позволяющие программе дифференцировать произносимые / непроизносимые согласные (гласные) в зависимости от их позиции и сочетаемости. Для каждого знака указывается общее количество его употреблений в предлагаемом тексте с той целью, чтобы восприятие аллитерационных и ассонансных повторений звуков было более наглядным при работе с большими фрагментами текста:



Отражение повторяющихся гласных звуков основано на понимании ассонанса как повтора не только ударных, но и безударных гласных звуков, что связано с все возрастающим интересом исследователей к ритму прозы, одним из средств ритмизации которого является ассонанс. Особенностью программы в данном аспекте является выделение в качестве ассонирующего не каждого звука в отдельности, а группы звуков, имеющих фонетическое сходство. Так, например, под транскрипционным знаком [a] подразумевается группа звуков [a, ã, wa, ua, ya, aj, ja], под транскрипционным знаком [e] объединены [ej, je, ɛ, jɛ, ɛj, yɛ, uɛ] и т. д.

В процессе создания программы также разрабатывается алгоритм анализа ритма на лексическом и грамматическом уровнях его реализации. Основными средствами, представленными в рамках данных уровней, являются лексические и грамматические повторы различных типов, а именно: деривация, анафора, симплока, эпифора, анадиплозис, эпаналепсис, редупликация, градационный повтор, однородные члены предложения, последовательность вопросительных, восклицательных предложений и апозиопеза.

Кроме того, программа предполагает поиск «ключевых слов» произведения. После ввода текста во всем его объеме в текстовое поле программой осуществляется поиск наиболее частотных лексем, которые в зависимости от значения могут рассматриваться исследователем как ключевые для введенного текста. Данный аспект был протестирован на материале романов О. де Бальзака, Г. Флобера, Г. де Мопассана. В результате была выявлена наибольшая частотность употребления имен собственных (имен героев произведений), а также лексем homme, femme, madame, monsieur и некоторых других. В выявлении ключевых слов произведения данные лексемы не принимаются во внимание в связи со слабой степенью отражения ими конкретной идеи текста. Данные лексемы являются частотными для большинства исследованных текстов и не отражают их специфики. Бóльшую важность имеют лексемы, употребленные с несколько меньшей частотностью. Именно они могут быть рассмотрены как ключевые. В этом случае исследователем должна осуществляться работа по отбору наиболее частотных лексем в зависимости от содержания текста. Так, для романов указанных выше авторов характерны повторы существительных amour, vie, monde, coeur, jeune, âme, argent, franc, mort и некоторые другие. Безусловно, состав данных лексем варьируется от произведения к произведению. Данный аспект анализа в определенной степени необходим для формирования ключевой идеи текста, для определения настроений и характеров, раскрываемых автором.

Таким образом, анализ ритма при помощи программы проводится на фонетическом, морфологическом и синтаксическом языковых уровнях, а в рамках литературоведческой стилистики выражается на фоностилистическом, лексическом, грамматическом, а также композиционном уровнях его реализации.

Создание программы для анализа ритма текста, прежде всего, необходимо для облегчения работы исследователя по таким параметрам, как подсчет слогов для выявления их равенства или последовательностей, имеющих ритмический характер, для механического выявления повторов, создающих ритмику текста и проявляющихся на всех языковых уровнях: фонетическом, лексическом и грамматическом. Все это необходимо для достижения высокой степени продуктивности в работе с текстом, проявляющейся в экономии времени исследователя, в исключении человеческого фактора, допускающего разного рода ошибки.

Следует отметить, что программа не является узкоспециализированной, направленной только на анализ художественных прозаических текстов. Предлагаемый комплекс средств и алгоритм их анализа могут быть применимы как к поэтическим тек-

стам, так и к прозаическим в рамках других функциональных стилей, например, научного, делового и др.

Компьютерные методы исследования текста все активнее проникают в лингвистику и филологию, заменяя в некоторых случаях человека. Для этого явления существует простое объяснение: во-первых, компьютерные программы, созданные для произведения определенных манипуляций с текстом, экономят время исследователя, во-вторых, они исключают погрешность, связанную с человеческим фактором, в-третьих, позволяют анализировать большие объемы материала. Предлагаемая программа анализа ритма текста ориентирована не только на лингвистов и филологов, занимающихся исследованием текста, но и на простых школьников, которые могут опираться на результаты работы программы при анализе поэтических текстов. Отметим, что компьютерная обработка материала не исключает мыслительного процесса исследователя, она лишь заменяет некоторые механические операции с целью экономии времени и сил человека.

Основными методологическими целями при работе с данной компьютерной программой в ходе преподавания иностранного языка является формирование умений и навыков работы с аутентичным текстом, формирование коммуникативной и информационной компетентности, а также стимулирование исследовательской деятельности учащихся.

Основными этапами работы с программой в ходе практических занятий по французскому языку (в частности, по анализу текста) могут служить: 1) анализ литературного уровня произведения, в ходе которого учащийся выражает свое мнение об основной идее текста, формулируемой на основании результатов поиска программой ключевых слов отрывка или всего произведения; 2) анализ речевого уровня, подкрепляемого программным поиском средств выразительности речи персонажей или авторского повествования. Данный подход в вопросе анализа текста в большей степени стимулирует учащихся к чтению аутентичной литературы, повышает уровень самообразовательных навыков, формирует умение ориентироваться в большом потоке информации, анализировать, сравнивать, аргументировать, обобщать и делать выводы.

Как уже было отмечено, материал статьи отражает начальный этап работы над созданием программы. Последующие изменения в программе, позволяющие анализировать ритмику текста с точки зрения различных языковых аспектов, будут отражены в цикле статей, посвященных данному вопросу.

СПИСОК ЛИТЕРАТУРЫ:

1. *Альфред В. Ахо, Моника С. Лам, Рави Сети, Джеффри Д. Ульман.* Компиляторы: принципы, технологии и инструментарий = Compilers: Principles, Techniques, and Tools [Электронный ресурс]. 2-е изд. М.: Вильямс, 2008. URL: http://dragonbook. stanford. edu/, дата обращения: 10.10.213.
2. *Бойчук Е. И.* Ритм прозы (на материале французского языка): Монография. Ярославль: Изд-во ЯГПУ, 2013. 224 с.

REFERENCES

1. *Al'fred V. Aho, Monika S. Lam, Ravi Seti, Dzheffri D.* Ul'man. Kompiljatory: printsipy, tehnologii i instrumentarij = Compilers: Principles, Techniques, and Tools [Elektronnyj resurs]. 2-e izd. M.: Vil'jams, 2008. URL: http://dragonbook. stanford. edu/, data obrashchenija: 10.10.213.
2. *Bojchuk E. I.* Ritm prozy (na materiale frantsuzskogo jazyka). Monografija: Jaroslavl': Izd-vo JaGPU, 2013. 224 s.