

Метод нечеткого анализа динамики стилей в стихотворных текстах

С.Н. Андреев¹, В.В. Борисов¹, К.П. Коршунова¹

¹ ФБГОУ ВО «Смоленский государственный университет» г. Смоленск

¹ Филиал ФБГОУ ВО «НИУ «МЭИ» г. Смоленск

Аннотация. Предлагается метод нечеткого анализа динамики изменения авторского стиля в стихотворных текстах. В основе метода лежит определение наиболее типичных (ядра) и отклоняющихся от канона (периферии) стихотворных произведений, а также анализ динамики изменения ядра/периферии этих произведений для различных творческих периодов. При этом система показателей, характеризующих близость текстов к центрам ядер, имеет сложную иерархическую структуру, базируется на подмножествах различных признаков (морфологических, синтаксических, ритмических) и позволяет учитывать уровни их совместимости. Применение метода проиллюстрировано на материале лирики М.Ю. Лермонтова.

Ключевые слова: нечеткий анализ, иерархическая система показателей, нечеткая оценочная модель, динамика стиля.

DOI 10.14357/20718594180318

Введение

В настоящее время наблюдается ярко выраженная тенденция к внесению «точности» в лингвистические исследования. Это достигается отказом от умозрительных выводов, построенных на рассмотрении отдельных примеров, и использованием методологии, основанной на следующих принципах: привлечение к анализу больших объемов текстов; их описание (разметка) единым для данного исследования списком признаков; поиск тенденций и закономерностей в организации речи на базе количественного анализа частотности языковых единиц, а также анализа соотношений признаков языковых единиц. Диапазон решаемых при этом задач весьма широк и включает в себя установление авторства анонимных и псевдонимных текстов, выявление гендерных различий речи, классификацию жанров, авторских стилей и самих авторов, выявление законов распределения элементов текста и др. [1–6].

Одним из основных направлений в рамках такого количественного подхода в лингвистике является исследование возможных изменений индивидуального стиля путем изучения варьирования разноуровневых и разноаспектных лингвистических признаков текстов в различные периоды творчества их авторов. Выявление вариативности стиля во времени с установлением интегральных (обеспечивающих преемственность стиля) и дифференциальных (варьирующихся во времени) признаков, позволяет решать такие задачи, как датировка текстов, периодизация авторского творчества, описание речи различных возрастных групп, создания текста, психические заболевания и характер их протекания, тенденции развития языковой личности [7–12], а также имеют большое значение при решении вопросов авторства [13].

Изменения стиля в наиболее эксплицитном виде наблюдаются в стихотворной речи, так как стихотворный текст обладает намного более строгой упорядоченностью, чем текст про-

✉ Борисов Вадим Владимирович e-mail: vbor67@mail.ru

заический – не только горизонтальной, но и вертикальной, поскольку включает в себя повторяющиеся и сопоставимые между собой речевые отрезки (стихотворные строки). Еще одним важным фактором организации стиха являются ограничения, накладываемые на выбор морфологических и фонетических классов слов в связи с требованиями метрической системы и ритма [14].

Следует отметить, что исследование языковых признаков стихотворного текста получило достаточно большое развитие в направлении «лингвистики стиха», в результате чего были выявлены закономерности распределения частей речи в различных позициях в стихе, распределения в стихе синтаксических связей различной силы, выбора слов определенной семантики и формальных признаков в первую (зачин стиха) и последнюю (рифма) ударные позиции, были проанализированы соотношения групп признаков стихотворной речи (ритмики, морфологии, рифменной системы, фонетики и др.) [14–19].

Изучению динамики стиля в стихотворной речи посвящены исследования, проводимые, в основном, на материале английского языка [20–23]. Указанные исследования строятся на высоком уровне обобщения хронологических классов признаков. В то же время возникает необходимость более детальной характеристики таких важных для рассмотрения динамики стиля аспектов, как система признаков в различные периоды творчества, характеризующая наиболее типичные для данного этапа произведения (ядро) и произведения, отклоняющиеся от канона и образующие периферию класса произведений данного творческого периода.

В статье предлагается метод нечеткого анализа динамики изменения авторского стиля в стихотворных текстах по периодам творчества. В основе метода лежит определение наиболее типичных (ядра) и отклоняющихся от канона (периферии) стихотворных произведений, а также анализ динамики изменения ядер/периферии этих произведений для различных творческих периодов. Применение метода проиллюстрировано на материале лирики М.Ю. Лермонтова.

1. Постановка задачи

Разделим множество признаков стихотворных текстов на подмножества:

$P = \{P_j\}$, $P_j = \{p_{m_j}\}$ – j -е подмножество признаков, $m_j = 1, \dots, M_j$, $j = 1, \dots, J$.

Из множества признаков выделим подмножества морфологических, синтаксических и ритмических признаков, являющихся релевантными для анализа стихотворного текста и использующихся для решения различных задач в стилеметрии [8, 16, 24–27].

Морфологические признаки учитывают соотношение частей речи: существительных, глаголов, прилагательных, наречий, определений-местоимений и местоимений:

$$P_1 = \{p_{m_1}\}, m_1 = 1, \dots, M_1, M_1 = 18.$$

Подсчет частей речи ведется как с учетом, так и без учета локализации слов. В первом случае анализируются слова, (а) замещающие первую сильную позицию ($p_{1_1} - p_{6_1}$ – признаки, определяющие число, соответственно, существительных, глаголов, прилагательных, наречий, определений-местоимений и местоимений), (б) замещающие последнюю сильную позицию в строке (признаки $p_{7_1} - p_{12_1}$). Во втором случае учитываются все слова данной части речи независимо от позиции, т.е. все существительные, глаголы, прилагательные, наречия, определения-местоимения и местоимения (признаки $p_{13_1} - p_{18_1}$).

Синтаксические признаки отражают специфику стихотворного текста:

$$P_2 = \{p_{m_2}\}, m_2 = 1, \dots, M_2, M_2 = 8,$$

где p_{1_2} – характеризует наличие паузы в стихотворной строке, вызванной наличием в строке двух предложений; p_{2_2} – определяет синтаксический (стихотворный) перенос, т.е. несовпадение синтаксического членения текста на синтагмы и стихового членения текста на строки, из-за чего синтаксическая пауза переходит с конца стиха на середину первой и/или второй из образующих перенос строк; p_{3_2} – строка, маркированная восклицательным, вопросительным знаком или многоточием; p_{4_2} – предложения, заканчивающиеся внутри строки и маркированные вопросительным, восклицательным знаком или многоточием; p_{5_2} – полная инверсия, т.е. инверсия подлежащего и сказуемого; p_{6_2} – бессоюзное предложение; p_{7_2} –

сочинительное предложение; p_{8_2} – подчинительное предложение.

Ритмические признаки:

$$P_3 = \{p_{m_3}\}, m_3 = 1, \dots, M_3, M_3 = 5,$$

где p_{1_3} – ударение на первом нечетном слоге – анакрусе; p_{2_3} – пропуск ударения на первом четном слоге; p_{3_3} – пропуск ударения на втором четном слоге; p_{4_3} – пропуск ударения на третьем четном слоге; p_{5_3} – полноударная строка, т.е. на все четные слоги строки падает ударение.

К анализу привлечено 80 лирических стихотворений М.Ю. Лермонтова, написанных четырехстопным ямбом и относящихся к различным периодам его творчества: $A = \{A_i\}$, где

$A_i = \{a_{n_i}\}$ – подмножество стихотворений, относящихся к i -му периоду творчества, $n_i = 1, \dots, N_i$, $i = 1, \dots, I$.

В данной работе применяется устоявшаяся трехчастная периодизация творчества М.Ю. Лермонтова, в соответствии с которой выделяются следующие периоды: 1828–1831 гг., 1832–1936 гг., 1837–1841 гг. Распределение текстов по периодам:

1-й период: $A_1 = \{a_{n_1}\}$, $n_1 = 1, \dots, N_1$, $N_1 = 48$;

2-й период: $A_2 = \{a_{n_2}\}$, $n_2 = 1, \dots, N_2$, $N_2 = 16$;

3-й период: $A_3 = \{a_{n_3}\}$, $n_3 = 1, \dots, N_3$, $N_3 = 16$.

Анализируемые в данном исследовании произведения М.Ю. Лермонтова представлены в Табл. 1.

Табл. 1. Анализируемые произведения М.Ю. Лермонтова

Обозначение и название произведения		Обозначение и название произведения		Обозначение и название произведения	
a_{1_1}	Поэт	a_{28_1}	Посвящение	a_{7_2}	Посвящение («Прими, прими мой грустный труд...»)
a_{2_1}	К Д...ву	a_{29_1}	«Вблизи тебя до этих пор...»	a_{8_2}	Парус
a_{3_1}	Пир	a_{30_1}	Нищий	a_{9_2}	«Слова разлуки повторяя...»
a_{4_1}	К... («Не привлекай меня красой!...»)	a_{31_1}	Опасение	a_{10_2}	Солнце
a_{5_1}	Мой демон	a_{32_1}	«Ты помнишь ли, как мы с тобою...»	a_{11_2}	«Ты молод. Цвет твоих кудрей...»
a_{6_1}	К*** («Глядися чаще в зеркала...»)	a_{33_1}	1831-го января	a_{12_2}	«В альбом Н.Ф. Ивановой»
a_{7_1}	Молитва («Не обвиняй меня, всесильный...»)	a_{34_1}	Звуки и взор	a_{13_2}	«Когда последнее мгновенье...»
a_{8_1}	Черкешенка	a_{35_1}	Романс к И... («Когда я унесу в чужбину...»)	a_{14_2}	«Я жить хочу! хочу печали...»
a_{9_1}	Война	a_{36_1}	«Когда б в покорности незнания...»	a_{15_2}	«Как луч зари, как розы Леля...»
a_{10_1}	К*** («Мы снова встретились с тобой»)	a_{37_1}	К кн. Л. Г-ой	a_{16_2}	«По произволу дивной власти»
a_{11_1}	Романс («Коварной жизнью недовольный...»)	a_{38_1}	«Пусть я кого-нибудь люблю...»	a_{1_3}	Ветка Палестины
a_{12_1}	Ответ	a_{39_1}	«Пора уснуть последним сном...»	a_{2_3}	«Расстались мы, но твой портрет...»
a_{13_1}	Баллада	a_{40_1}	К другу В. Ш<еншину>	a_{3_3}	«Я не хочу, чтоб свет узнал...»
a_{14_1}	К* (Из Шиллера)	a_{41_1}	«Как дух отчаянья и зла...»	a_{4_3}	«Спеша на север из далека...»
a_{15_1}	К Грузинову	a_{42_1}	«Я не для ангелов и рая...»	a_{5_3}	Она поет
a_{16_1}	Жена севера	a_{43_1}	«Метель шумит и снег валит...»	a_{6_3}	«Гляжу на будущность с боязнью...»
a_{17_1}	Раскаянье	a_{44_1}	Силуэт	a_{7_3}	А.Г. Хомутовой
a_{18_1}	К*** («Не говори: одним высоким...»)	a_{45_1}	23-го марта 1831 г.	a_{8_3}	Вид гор из степей Козлова

продолжение таблицы

Обозначение и название произведения		Обозначение и название произведения		Обозначение и название произведения	
a_{19_1}	В альбом («Нет! — я не требую вниманья...»)	a_{46_1}	«Я не люблю тебя; страстей...»	a_{9_3}	А.О. Смирновой
a_{20_1}	Кавказу	a_{47_1}	«Зови надежду - сновиденьем...»	a_{10_3}	М. П. Соломирской
a_{21_1}	11 июля	a_{48_1}	Трубецкому (Новогодние эпиграммы)	a_{11_3}	Из альбома С.Н. Карамзиной
a_{22_1}	Чума в Саратове («Чума явилась в наш предел...»)	a_{1_2}	«Люблю я цепи синих гор...»	a_{12_3}	Графине Ростопчиной
a_{23_1}	Стансы («Взгляни, как мой спокоен взор...»)	a_{2_2}	Прелестнице	a_{13_3}	Договор
a_{24_1}	К*** («Когда к тебе молвы рассказ...»)	a_{3_2}	«Нет, я не Байрон, я другой...»	a_{14_3}	«Прощай, немытая Россия...»
a_{25_1}	«Свершилось! Полно ожидать...»	a_{4_2}	К* («Оставь напрасные заботы...»)	a_{15_3}	Пророк
a_{26_1}	(Глупой красавице)	a_{5_2}	К* («Мой друг, напрасное старанье!...»)	a_{16_3}	Оправдание
a_{27_1}	«Нередко люди и бранили...»	a_{6_2}	«Она не гордой красотою...»		

Значения всех признаков из множеств P_1 , P_2 и P_3 , характеризующих эти стихотворные тексты, нормированы относительно числа строк каждого произведения.

Предлагаемый метод должен обеспечивать:

- определение типичных и отклоняющихся от канона стихотворных произведений для различных творческих периодов относительно как отдельных подмножеств, так и всего множества признаков;
- определение степени принадлежности стихотворных произведений к центру ядра с учетом иерархической структуры оценивания и различного уровня совместимости подмножеств показателей;
- анализ «плотности»/«разреженности» текстов внутри и вокруг ядер относительно отдельных подмножеств и всего множества признаков;
- анализ динамики изменения ядер/периферии стихотворных произведений по различным творческим периодам;
- исследование тенденций изменения отдельных показателей, подмножеств показателей, характеристик ядер и периферии стихотворных произведений по различным периодам творчества.

2. Описание метода и результаты анализа динамики стилей в стихотворных текстах

Предлагаемый метод включает в себя рассмотренные ниже этапы.

Этап 1: определение центров ядер стихотворных произведений (объектов) отдельно для подмножеств морфологических, синтаксических и ритмических признаков по всем периодам творчества. В качестве процедуры определения центра ядра объектов используем предлагаемую разновидность возможностного алгоритма нечеткой кластеризации (PFCM – Possibilistic Fuzzy C-Means Algorithm [28, 29]), применяемого для случая единственного кластера.

Примечание: для краткой иллюстрации предлагаемой разновидности алгоритма PFCM воспользуемся упрощенными обозначениями объектов и координат центров ядер.

Алгоритм PFCM для данного случая сводится к итеративной процедуре определения координат $\{c_1, c_2, \dots, c_M\}$ (где M – число признаков $\{p_1, p_2, \dots, p_M\}$) центра ядра рассматриваемых объектов $A = \{a_1, a_2, \dots, a_N\}$ за счет минимизации целевой функции:

$$E(u_n, c) = \sum_{n=1}^N u^\beta(a_n) d^2(c, p(a_n)) + \eta \sum_{n=1}^N (1 - u(a_n))^\beta,$$

при ограничении

$$u(a_n) > 0, \forall n \in \{1, \dots, N\},$$

где $u(a_n) = \{u_1(a_n), u_2(a_n), \dots, u_M(a_n)\}$ иницируются на начальном шаге алгоритма случайными числами; β – параметр (обычно $\beta = 2$), характеризующий степень нечеткости итогового распределения; $d(c, p(a_n))$ – расстояние между текущим центром ядра c и объектом a_n в принятой метрике; η – параметр, определяющий расстояние до центра кластера, на котором значение степени принадлежности к кластеру становится равным 0,5. Параметр η вычисляется следующим образом:

$$\eta = \frac{\sum_{n=1}^N u^\beta(a_n) d^2(c, p(a_n))}{\sum_{n=1}^N u^\beta(a_n)}.$$

Таким образом, по результатам данного этапа определяется совокупность C_{A_i, P_j} центров ядер $c_{A_i, P_{m_j}}$ подмножеств объектов по всем периодам творчества относительно каждого из морфологических, синтаксических и ритмических признаков:

$$C_{A_i, P_j} = \{c_{A_i, P_{m_j}}\}, i = 1, \dots, I, m_j = 1, \dots, M_j, j = 1, \dots, J,$$

где $c_{A_i, P_{m_j}}$ – центр ядра подмножества объектов

$$A_i = \{a_{n_i}\} \text{ относительно признака } p_{m_j}.$$

Этап 2: определение степени близости подмножеств объектов для всех периодов творчества к центрам ядер для каждого подмножества признаков.

В качестве показателя близости объекта a_{n_i} i -го периода к центру ядра относительно j -го подмножества признаков целесообразно использовать евклидово расстояние:

$$d(C_{A_i, P_j}, P_j(a_{n_i})) = \frac{1}{\sqrt{M_j}} \sqrt{\sum_{m_j=1}^{M_j} (c_{A_i, P_{m_j}} - p_{m_j}(a_{n_i}))^2},$$

$$j = 1, \dots, J, n_i = 1, \dots, N_i, i = 1, \dots, I,$$

где $P_j(a_{n_i}) = \{p_{m_j}(a_{n_i})\}$ – совокупность значений признаков объекта a_{n_i} , $m_j = 1, \dots, M_j$.

Примечания:

1. В дальнейшем для наглядности вместо $d(C_{A_i, P_j}, P_j(a_{n_i}))$ будем использовать обозначение $d_j(a_{n_i})$.

2. В Табл. 3 (стр. 84) сведены вычисленные показатели близости объектов a_{n_i} ($n_i = 1, \dots, N_i, i = 1, \dots, I$) к центрам ядер относительно подмножеств морфологических, синтаксических и ритмических признаков по периодам творчества – $d_1(a_{n_i})$, $d_2(a_{n_i})$ и $d_3(a_{n_i})$, соответственно.

Этап 3: построение нечеткой оценочной модели с учетом совместимости подмножеств показателей близости объектов к центрам ядер.

Для построения данной оценочной модели используем способ, предложенный в [30], включающий в себя шаги.

Шаг 1. Построение структуры нечеткой модели с учетом уровней совместимости подмножеств показателей близости объектов к центрам ядер.

На Рис. 1 показана структура построенной нечеткой модели для обобщенной оценки степени близости объектов к центрам ядер относительно всего множества признаков для различных периодов творчества.

Шаг 2. Определение уровней совместимости между показателями близости объектов к центрам ядер, относящимися к подмножествам морфологических, синтаксических и ритмических признаков.

На Рис. 1 показаны уровни совместимости con_{kl} ($k, l = 1, \dots, J$) между показателями $d_j(a_{n_i})$ ($j = 1, \dots, J$) близости объектов к центрам ядер, относящихся к подмножествам морфологических, синтаксических и ритмических признаков. При этом:

$$con_{kl} \in \{NC, LC, MC, HC, FC\}, k, l = 1, \dots, J,$$

где NC – низкий уровень совместимости, LC – уровень совместимости ниже среднего, MC – средний уровень совместимости, HC – уровень совместимости выше среднего, FC – высокий уровень совместимости.

Уровни совместимости con_{kl} ($k, l = 1, \dots, J$) между показателями $d_j(a_{n_i})$ ($j = 1, \dots, J$) близости объектов к центрам ядер, относящимися к подмножествам морфологических, синтаксических и ритмических признаков, могут трактоваться с точки зрения их корреляции или

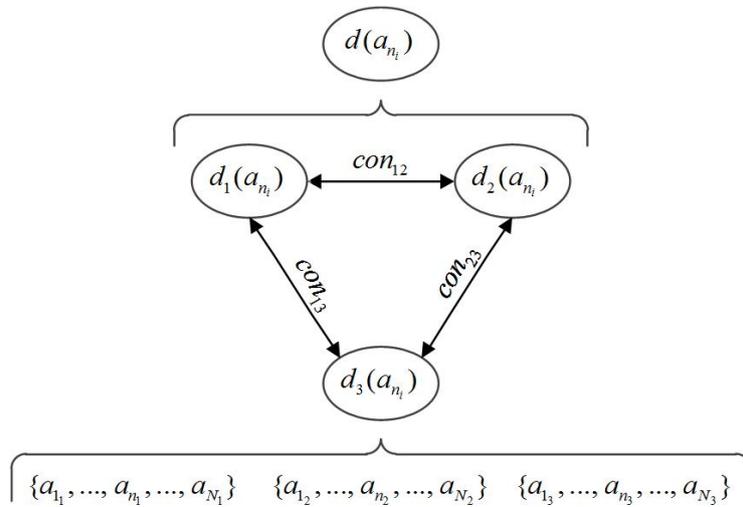


Рис. 1. Структура нечеткой модели для обобщенной оценки степени близости объектов к центрам ядер по всему множеству признаков

взаимовлияния и определяться как аналитически, так и экспертным путем.

В данном случае для определения уровней совместимости con_{kl} ($k, l = 1, \dots, J, k \neq l$) между соответствующими показателями $d_k(a_{n_i})$ и $d_l(a_{n_i})$ использовался подход, основанный на формировании и анализе так называемых «матриц сочетаемости», предложенный в [31]. В результате определены следующие уровни совместимости между этими показателями:

$$con_{12} = LC; con_{13} = LC; con_{23} = NC.$$

Шаг 3. Сопоставление уровней совместимости показателей близости объектов к центрам ядер с операциями их свертки.

В качестве операций свертки агрегируемых показателей выбраны операции \min и \max , а также параметризованное семейство операций свертки типа:

$$\text{med}(d_k(a_{n_i}), d_l(a_{n_i}); \alpha), k, l \in \{1, \dots, J\}, \alpha \in [0, 1],$$

где α – параметр, характеризующий степень совместимости показателей, относящихся к различным подмножествам признаков.

В Табл. 2 представлено сопоставление уровней совместимости агрегируемых показателей с операциями их свертки.

Шаг 4. Задание стратегии оценивания и формирование обобщенной оценки степени близости объектов к центру ядра относительно всего множества признаков.

Табл. 2. Сопоставление уровней совместимости показателей с операциями их свертки

№ п.п	Уровень совместимости показателей	Операция свертки показателей $d_k(a_{n_i})$ и $d_l(a_{n_i})$
1	NC	$\min(d_k(a_{n_i}), d_l(a_{n_i}))$
2	LC	$\text{med}(d_k(a_{n_i}), d_l(a_{n_i}); 0,25)$
3	MC	$\text{med}(d_k(a_{n_i}), d_l(a_{n_i}); 0,5)$
4	HC	$\text{med}(d_k(a_{n_i}), d_l(a_{n_i}); 0,75)$
5	FC	$\max(d_k(a_{n_i}), d_l(a_{n_i}))$

Стратегия оценивания определяет порядок просмотра уровней совместимости показателей близости объектов к центрам ядер, обуславливающий очередность их агрегирования в модели с использованием соответствующих операций свертки, а также пересмотр уровней совместимости после каждой операции свертки показателей.

Могут быть заданы следующие стратегии оценивания: от наиболее совместимых показателей к наименее совместимым, от наименее совместимых к наиболее совместимым. В данном исследовании используется стратегия оценивания от наиболее совместимых к наименее совместимым показателям.

В результате применения рассмотренного способа сформирована модель, позволяющая получить обобщенную оценку степени близости

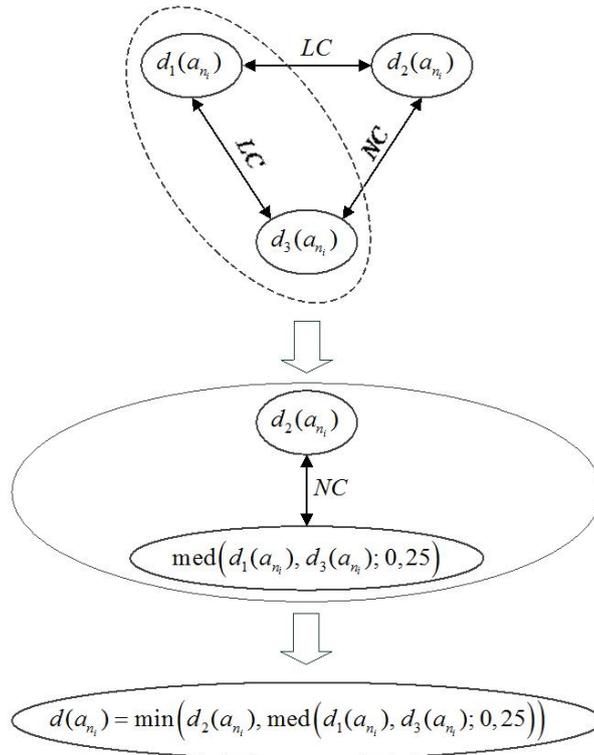


Рис. 2. Иллюстрация процедуры получения обобщенной оценки степени близости объектов к центру ядра относительно всего множества признаков

сти объектов к центру ядра относительно всего множества признаков в соответствии со следующим выражением:

$$d(a_{n_i}) = \min(d_2(a_{n_i}), \text{med}(d_1(a_{n_i}), d_3(a_{n_i}); 0,25)).$$

На Рис. 2 показана процедура получения обобщенной оценки степени близости объектов

к центру ядра относительно всего множества признаков с использованием этой модели.

В Табл. 3 представлены полученные результаты оценки близости объектов к центрам ядер как относительно подмножеств морфологических, синтаксических и ритмических признаков, так и относительно всего множества признаков в целом по периодам творчества М.Ю. Лермонтова.

Табл. 3. Показатели близости стихотворных текстов к центрам ядер относительно подмножеств и всего множества признаков по периодам творчества М.Ю. Лермонтова

Периоды творчества	Обозначение произведения	Показатели близости к центрам ядер				Периоды творчества	Обозначение произведения	Показатели близости к центрам ядер				Периоды творчества	Обозначение произведения	Показатели близости к центрам ядер			
		по подмножеству морфологических признаков	по подмножеству синтаксических признаков	по подмножеству ритмических признаков	по всему множеству признаков			по подмножеству морфологических признаков	по подмножеству синтаксических признаков	по подмножеству ритмических признаков	по всему множеству признаков			по подмножеству морфологических признаков	по подмножеству синтаксических признаков	по подмножеству ритмических признаков	по всему множеству признаков
		$d_1(a_{n_i})$	$d_2(a_{n_i})$	$d_3(a_{n_i})$	$d(a_{n_i})$			$d_1(a_{n_i})$	$d_2(a_{n_i})$	$d_3(a_{n_i})$	$d(a_{n_i})$			$d_1(a_{n_i})$	$d_2(a_{n_i})$	$d_3(a_{n_i})$	$d(a_{n_i})$
1	a_{1_i}	0,42	0,85	0,94	0,42	1	a_{28_i}	0,42	0,83	0,81	0,42	2	a_{7_i}	0,49	0,53	0,79	0,49
	a_{2_i}	0,45	0,74	0,95	0,45		a_{29_i}	0,60	0,74	0,37	0,60		a_{8_i}	0,34	0,34	0,31	0,34
	a_{3_i}	0,40	0,61	0,78	0,40		a_{30_i}	0,58	0,77	0,58	0,58		a_{9_i}	0,59	0,31	0,78	0,31

продолжение таблицы

a_{4_1}	0,63	0,29	0,54	0,29	2	a_{31_1}	0,79	0,71	0,68	0,71	3	a_{10_2}	0,30	0,50	0,37	0,30
a_{5_1}	0,66	0,71	0,86	0,66		a_{32_1}	0,49	0,64	0,67	0,49		a_{11_2}	0,71	0,78	0,73	0,71
a_{6_1}	0,67	0,79	0,61	0,67		a_{33_1}	0,72	0,49	0,80	0,49		a_{12_2}	0,40	0,46	0,61	0,40
a_{7_1}	0,85	0,56	0,71	0,56		a_{34_1}	0,63	0,70	0,51	0,63		a_{13_2}	0,42	0,44	0,45	0,42
a_{8_1}	0,51	0,73	0,84	0,51		a_{35_1}	0,56	0,42	0,56	0,42		a_{14_2}	0,40	0,53	0,88	0,40
a_{9_1}	0,42	0,66	0,77	0,42		a_{36_1}	0,73	0,56	0,74	0,56		a_{15_2}	0,33	0,75	0,71	0,33
a_{10_1}	0,43	0,47	0,61	0,43		a_{37_1}	0,69	0,70	0,55	0,69		a_{16_2}	0,44	0,65	0,57	0,44
a_{11_1}	0,48	0,50	0,79	0,48		a_{38_1}	0,41	0,54	0,55	0,41		a_{1_3}	0,69	0,55	0,75	0,55
a_{12_1}	0,62	0,58	0,60	0,58		a_{39_1}	0,25	0,43	0,53	0,25		a_{2_3}	0,45	0,71	0,86	0,45
a_{13_1}	0,55	0,75	0,33	0,55		a_{40_1}	0,53	0,63	0,85	0,53		a_{3_3}	0,47	0,38	0,64	0,38
a_{14_1}	0,26	0,49	0,41	0,26		a_{41_1}	0,28	0,64	0,50	0,28		a_{4_3}	0,83	0,80	0,86	0,80
a_{15_1}	0,37	0,31	0,43	0,31		a_{42_1}	0,58	0,83	0,58	0,58		a_{5_3}	0,43	0,34	0,31	0,34
a_{16_1}	0,61	0,78	0,69	0,61		a_{43_1}	0,49	0,65	0,83	0,49		a_{6_3}	0,55	0,53	0,72	0,53
a_{17_1}	0,84	0,81	0,68	0,81		a_{44_1}	0,31	0,36	0,74	0,31		a_{7_3}	0,65	0,71	0,64	0,65
a_{18_1}	0,47	0,58	0,47	0,47		a_{45_1}	0,48	0,68	0,66	0,48		a_{8_3}	0,58	0,63	0,85	0,58
a_{19_1}	0,52	0,59	0,47	0,52		a_{46_1}	0,42	0,43	0,74	0,42		a_{9_3}	0,22	0,44	0,66	0,25
a_{20_1}	0,59	0,48	0,59	0,48		a_{47_1}	0,40	0,36	0,59	0,36		a_{10_3}	0,47	0,60	0,51	0,47
a_{21_1}	0,73	0,85	0,68	0,73		a_{48_1}	0,37	0,36	0,64	0,36		a_{11_3}	0,42	0,68	0,83	0,42
a_{22_1}	0,63	0,73	0,88	0,63		a_{1_2}	0,79	0,83	0,57	0,79		a_{12_3}	0,47	0,73	0,91	0,47
a_{23_1}	0,46	0,73	0,53	0,46	a_{2_2}	0,49	0,83	0,67	0,49	a_{13_3}	0,51	0,68	0,94	0,51		
a_{24_1}	0,41	0,56	0,90	0,41	a_{3_2}	0,42	0,51	0,51	0,42	a_{14_3}	0,28	0,43	0,35	0,28		
a_{25_1}	0,55	0,27	0,71	0,27	a_{4_2}	0,69	0,60	0,68	0,60	a_{15_3}	0,44	0,78	0,93	0,44		
a_{26_1}	0,30	0,77	0,49	0,30	a_{5_2}	0,61	0,69	0,77	0,61	a_{16_3}	0,63	0,36	0,67	0,36		
a_{27_1}	0,38	0,88	0,56	0,38	a_{6_2}	0,59	0,60	0,54	0,59							

Этап 4: выделение ядер/периферии объектов для различных творческих периодов.

Для выделения ядер и периферии объектов для отдельного творческого периода можно непосредственно задать граничные значения $d(a_{n_i})$ отнесения объекта к ядру и периферии (например, $core_i = 0,8$ и $per_i = 0,2$).

В данной работе для решения задачи определения граничных значений ядер и периферии предлагается воспользоваться результатами анализа процедуры построения зависимости количества объектов (с накоплением) от степени близости к центру ядра. При этом для определения этих граничных значений предлагается воспользоваться критериями, наподобие критерия «каменистой осыпи». Так, в качестве нижнего граничного значения ядра $core_i$ произве-

дений предлагается выбрать такое значение $d(a_{n_i})$ на графике, для которого убывание количества объектов справа налево максимально ускоряется. А в качестве нижнего граничного значения периферии per_i можно выбрать значение $d(a_{n_i})$, для которого убывание числа объектов справа налево максимально замедляется. На Рис. 3–Рис. 5 показаны граничные значения ядер и периферии объектов для различных творческих периодов, выбранных в соответствии с этими критериями.

Таким образом, в соответствии с проведенным на основе предлагаемого метода анализом к ядрам относятся следующие стихотворные произведения (по степени близости к центру ядра):

- для первого периода: «Раскаянье», «11 июля», «Опасение»;

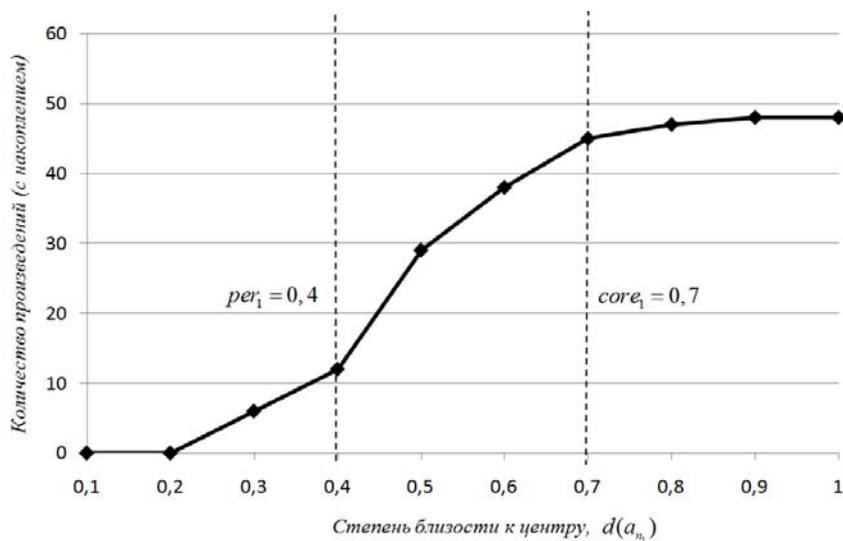


Рис. 3. Распределение стихотворных текстов первого периода по степени близости к центру ядра ($core_1 = 0,7$, $per_1 = 0,4$)

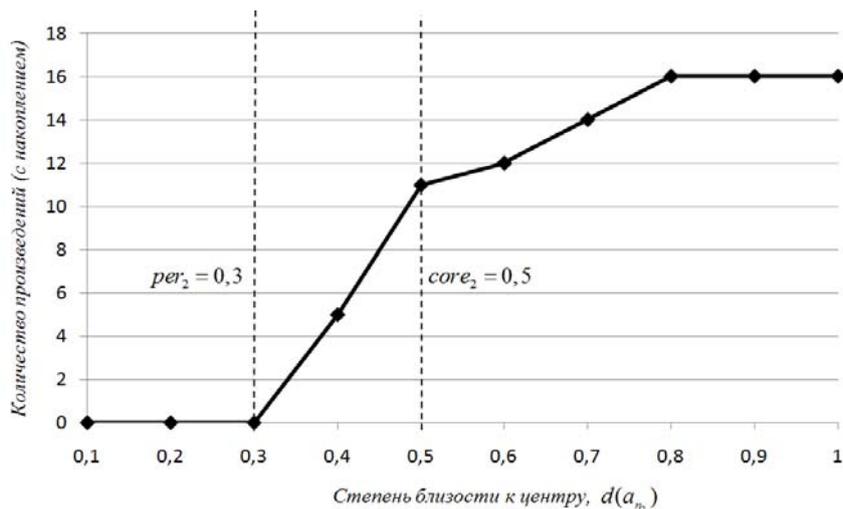


Рис. 4. Распределение стихотворных текстов второго периода по степени близости к центру ядра ($core_2 = 0,5$, $per_2 = 0,3$)

- для второго периода: «Люблю я цепи синих гор...», «Ты молод. Цвет твоих кудрей...», «К*» («Мой друг, напрасное старанье!...»), «К*» («Оставь напрасные заботы...»), «Она не гордой красотой...»;
- для третьего периода: «Спеша на север из далека», «А.Г. Хомутовой».

Примечание: нахождение ядер/периферии стихотворных произведений может быть выполнено как относительно всего множества признаков, так и относительно подмножеств морфологических, синтаксических и ритмических признаков.

Этап 5: анализ тенденций изменения отдельных показателей близости стихотворных текстов к центрам ядер относительно различных морфологических, синтаксических и ритмических признаков по различным периодам творчества.

На Рис. 6 показаны изменения координат центров ядер относительно различных морфологических, синтаксических и ритмических признаков по различным периодам творчества.

По результатам анализа этих тенденций можно сделать следующие выводы:

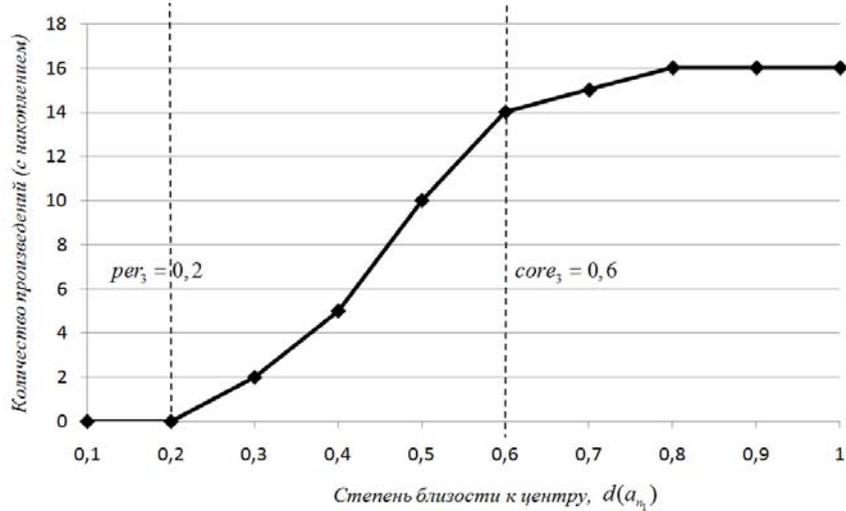


Рис. 5. Распределение стихотворных текстов третьего периода по степени близости к центру ядра ($core_2 = 0,6$, $per_2 = 0,2$)

- среди морфологических признаков наибольшей динамике оказались подвержены (по степени изменения): p_{21} , p_{8} , p_{13} , p_{17} ;
- среди синтаксических признаков наибольшая динамика свойственна (по степени изменения): p_{52} , p_{12} , p_{32} ;
- среди ритмических признаков наибольшая динамика свойственна (по степени изменения): p_{53} , p_{23} , p_{13} .

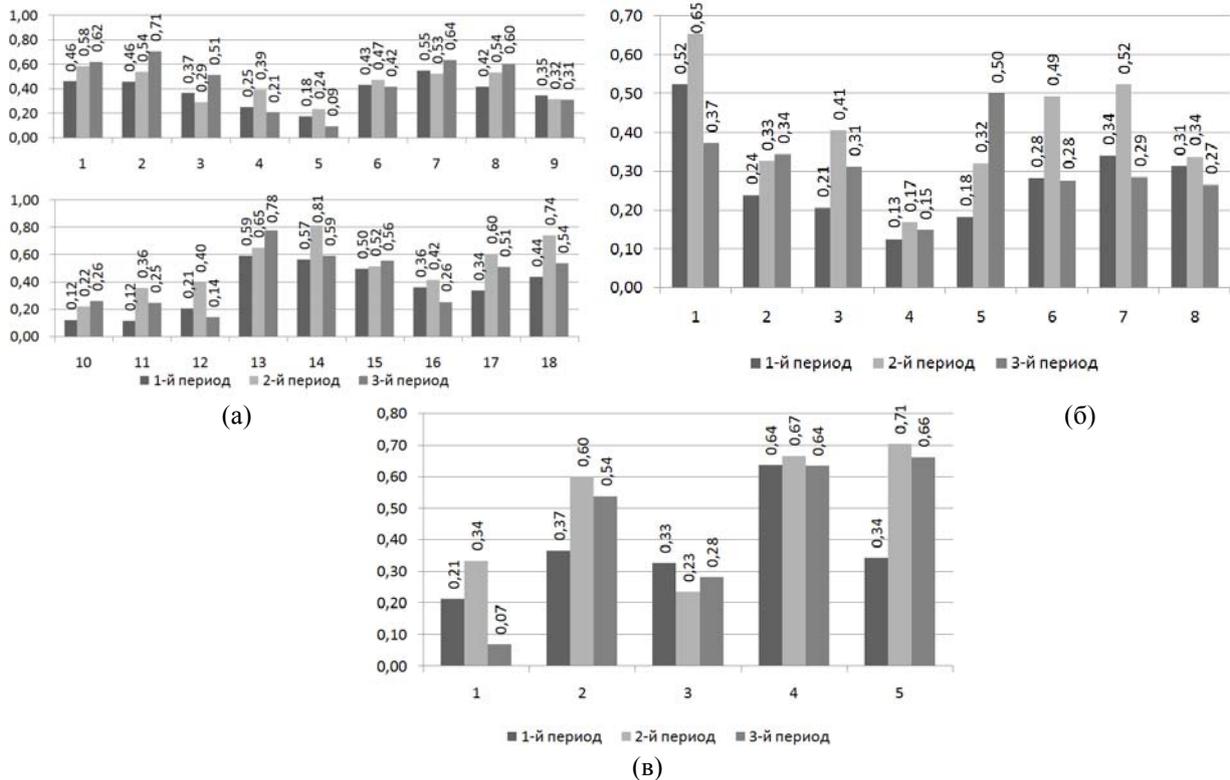


Рис. 6. Изменения координат центров ядер относительно морфологических $p_1 - p_{18}$ (а), синтаксических $p_{12} - p_{18}$ (б) и ритмических $p_{13} - p_{18}$ (в) признаков по трем периодам творчества М.Ю. Лермонтова

Заключение

В статье предлагается метод нечеткого анализа динамики изменения авторского стиля в стихотворных текстах по периодам творчества.

В основе метода лежит определение наиболее типичных (ядра) и отклоняющихся от канона (периферии) стихотворных произведений, а также анализ динамики изменения этих ядер и периферии для различных творческих периодов. При этом система показателей, характеризующих близость текстов к центрам ядер, имеет сложную иерархическую структуру, базируется на подмножествах морфологических, синтаксических и ритмических признаков и позволяет учитывать различные уровни их совместимости.

Предлагаемый метод позволяет осуществить постановку и решение задач анализа стиля в стихотворных текстах, включая:

1) определение типичных и отклоняющихся от канона произведений для различных творческих периодов как относительно отдельных подмножеств признаков, так и всего множества признаков;

2) определение степени близости произведений к центру ядра с учетом иерархической структуры оценивания и различного уровня совместимости подмножеств показателей;

3) анализ «плотности»/«разреженности» текстов внутри и вокруг ядер относительно отдельных подмножеств и всего множества признаков;

4) анализ динамики изменения ядер/периферии стихотворных произведений для различных творческих периодов;

5) исследование тенденций изменения отдельных показателей, подмножеств показателей, характеристик ядер и периферии стихотворных произведений по различным периодам творчества.

Применение предлагаемого метода проиллюстрировано на примере анализа динамики изменения авторского стиля в лирических произведениях М.Ю. Лермонтова по трем периодам его творчества.

Литература

- Марусенко М.А. Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов. Л.: Изд-во Ленинградского ун-та, 1990.
- Altmann G. Supra-sentence levels // *Glottotheory*. 2014. Vol. 5 (1). P. 25–39.
- Koppel M, Argamon S., Shimoni A.R. Automatically categorizing written texts by author gender // *Literary and Linguistic Computing*. 2002. Vol. 17. No. 4. P. 401–412.
- Holmes D. The Evolution of stylometry in humanities scholarship // *Literary and Linguistic Computing*. 1998. Vol. 13, No. 3, P. 111–117.
- Hoover D.L. Altered texts, altered worlds, altered styles // *Language and Literature*. 2004. Vol. 13 (2). P. 99–118.
- Rudman J. Authorship Attribution: Statistical and Computational Methods // *Encyclopedia of Language & Linguistics, Second Edition*. Oxford: Elsevier, 2006. Vol. 1. P. 611–617.
- Караулов Ю.Н. Русская языковая личность и задачи её изучения // *Язык и личность*. М., 1989. С. 3–8.
- Мартыненко Г.Я. Ритмико-смысловая динамика русского классического сонета. СПбГУ, 2004. 30 с.
- Пиотровский Р.Г. Лингвистическая синергетика: исходные положения, первые результаты, перспективы. Филологический факультет СПбПУ, 2006.
- Koppel M., Akiva N. and Dagan I. Feature instability as a criterion for selecting potential style markers // *Journal of the American Society for Information Science and Technology*. 2006. Vol. 57, Is. 11. P. 1519–1525.
- McMenamin G.R. Forensic stylistics: Advances in forensic stylistics / by Gerald R. Boca Raton. London, New York, Washington D.C.: CRC Press LLC, 2002.
- Pennebaker J.M., Stone L.D. Words of wisdom: language use over the lifespan // *Journal of Personality and Social Psychology*. 2003. Vol. 85. No. 2. P. 291–301.
- Hoover D.L. Stylometry, chronology, and the styles of Henry James // *Proceedings of Digital Humanities 2006*. Paris, 2006. P. 78–80.
- Гаспаров М.Л., Скулачева Т.В. Статьи о лингвистике стиха. М.: Языки славянской культуры, 2004.
- Гаспаров М.Л. Синтаксическая структура стихотворной строки /Материалы международной конференции «Славянский стих: Лингвистическая и прикладная поэтика». 1998. М.: Языки славянской культуры, 2001. С. 130–137.
- Гаспаров М.Л. Точные методы анализа грамматики в стихе / М.Л. Гаспаров. Избранные труды. 2012. Т. 4. С. 23–35.
- Гаспаров М.Л. Пушкин и проблемы поэтической формы: язык и стих // М.Л. Гаспаров. Избранные труды. 2012. Т. 4. С. 117–129.
- Шоу Дж.Т. Сомнительные «ё» в рифмах послелицейской поэзии Пушкина // *Славянский стих. Лингвистическая и прикладная поэтика*. М.: Языки славянской культуры, 2001. С. 267–274.
- Popescu I-I, Chech R, Best K-H., Altmann G. Descriptivity in Slovak lyrics // *Glottotheory*. 2013. V. 4 (1), P. 92–104.
- Андреев В.С. Квантитативное описание индивидуально-авторской метафорической системы Г. Лонгфелло // *Известия смоленского государственного университета*. 2015. № 3(31). С. 201–210.
- Andreev S.N., Borisov V.V. Linguistic Analysis Based on Fuzzy Similarity models/ In “Sequences in Language and Text”. Iss. 69. P. 7–34./ Edited by G.K. Mikros, J. Macutek. Berlin/Boston: De Gruyter Mouton, 2015.
- Can F., Patton J.M. Change of word characteristics in 20th-Century Turkish literature: A statistical analysis // *Journal of quantitative linguistics*. 2010. Vol. 17. No. 3. P. 167–190.

23. Juola P. Becoming Jack London // *Journal of Quantitative Linguistics*. 2007. Vol. 14. No. 2. P. 145–147.
24. Андреев В.С. Моделирование индивидуального стиля (на основе лингвистических характеристик). М.: Флинта–Наука, 2012.
25. Баевский В.С., Семенова Н.А. Эволюция лирического стиха как основа периодизации творческой биографии поэта: кластерный и корреляционный анализ (Пушкин, Гумилев, Пастернак) // *Славянский стих VII: Лингвистика и структура стиха* / под ред. М.Л. Гаспарова, Т.В. Скулачевой. М.: Языки славянской культуры, 2004. С. 421–436.
26. Красноперова М.А. Основы реконструктивного моделирования стихосложения: На материале ритмики русского стиха. Издательство С.-Петерб. ун-та, 2000.
27. Матяш С.А. Переносы (enjambements) в лирике А.С. Пушкина и М.Ю. Лермонтова // *Вестник Оренбургского государственного университета*. 2006. №11. С. 57–63.
28. Krishnapuram R., Keller J.M. A possibilistic approach to clustering // *IEEE Transactions on Fuzzy Systems*. 1993. Vol. 1. No. 2. P. 98–110.
29. Krishnapuram R., Keller J.M. The possibilistic C-Means Algorithm: Insights and Recommendations // *IEEE Transactions on Fuzzy Systems*. 1996. Vol. 4. No. 3. P. 385–393.
30. Борисов В.В., Андреев С.Н., Федулов Я.А. Анализ сложных лингвистических объектов на основе нечетких оценочных моделей // *Искусственный интеллект и принятие решений*. 2014. №3. С.95–107.
31. Борисов В.В., Зернов М.М., Федулов Я.А. Способ нечеткого многокритериального оценивания с учетом согласованности параметров оценки // *Информационный бюллетень Смоленского регионального отделения Академии военных наук*. Вып. 28. Смоленск: ВА войсковой ПВО ВС РФ, 2013. С. 122–135.

Fuzzy analysis method of style dynamics in poetic texts

S.N. Andreev^I, V.V. Borisov^{II}, K.P. Korshunova^{II}

^I Smolensk State University, Smolensk, Russia

^{II} The Branch of National Research University "Moscow Power Engineering Institute" in Smolensk, Smolensk, Russia

We propose the fuzzy analysis method of style dynamics in poetic texts. The method is based on determination of the most typical (the core) and the most remote from the canon (the periphery) poetic texts and analysis of the core/the periphery progression dynamics for poems of different periods. At the same time, the parameters that characterize the poems closeness to the respective core form a system with complex hierarchical structure. The system is based on the subsets of different kinds of features: morphological, syntactic and rhythmical, – and allows to take into account their compatibility. The proposed method application is based on the M.Y. Lermontov's poems analysis.

Keywords: fuzzy analysis, hierarchical system of characteristics, fuzzy estimation (similarity) model, style dynamics.

DOI 10.14357/20718594180318

References

1. Marusenko M.A. 1990. Atributsiya anonimnykh i psevdonimnykh literaturnykh proizvedeniy metodami teorii raspoznavaniya obrazov [Attribution of anonymous and pseudonymous literary works using methods of the theory of pattern recognition]. Leningrad: publishing office of Leningrad State University.
2. Altmann G. 2014. Supra-sentence levels // *Glottology*. Vol. 5 (1). P. 25–39.
3. Koppel M, Argamon S., Shimon A.R. 2002. Automatically categorizing written texts by author gender // *Literary and Linguistic Computing*. Vol. 17. No. 4. P. 401–412.
4. Holmes D. 1998. The Evolution of stylometry in humanities scholarship // *Literary and Linguistic Computing*. Vol. 13, No. 3, P. 111–117.
5. Hoover D.L. 2004. Altered texts, altered worlds, altered styles // *Language and Literature*. Vol. 13 (2). P. 99–118.
6. Rudman J. 2006. Authorship Attribution: Statistical and Computational Methods // *Encyclopedia of Language & Linguistics*, Second Edition. Oxford: Elsevier. Vol. 1. P. 611–617.
7. Karaulov YU.N. 1989. Russkaya yazykovaya lichnost' i zadachi yeyo izucheniya [Russian language personality and the tasks of its study]. *Yazyk i lichnost'* [Language and Personality]. Moscow: 3–8.
8. Martynenko G.YA. 2004. Ritmiko-smyslovaya dinamika russkogo klassicheskogo sonata [The rhythmic and semantic dynamics of the Russian classical sonnet]. St. Petersburg: St. Petersburg State University. 30 p.
9. Piotrovskiy R.G. 2006. Lingvisticheskaya sinergetika: iskhodnyye polozheniya, pervyye rezul'taty, perspektivy [Linguistic synergetics: initial positions, first results, prospects]. St. Petersburg: Philology Department of the St. Petersburg State University.
10. Koppel M., Akiva N. and Dagan I. 2006. Feature instability as a criterion for selecting potential style markers // *Journal of the American Society for Information Science and Technology*. Vol. 57, Is. 11. P. 1519–1525.

11. McMenamin G.R. 2002. Forensic stylistics: Advances in forensic stylistics / by Gerald R. Boca Raton. London, New York, Washington D.C.: CRC Press LLC.
12. Pennebaker J.M., Stone L.D. 2003. Words of wisdom: language use over the lifespan // *Journal of Personality and Social Psychology*. Vol. 85. No. 2. P. 291–301.
13. Hoover D.L. 2006. Stylometry, chronology, and the styles of Henry James // *Proceedings of Digital Humanities 2006*. Paris. P. 78–80.
14. Gasparov M.L., Skulacheva T.V. 2004. Stat'i o lingvistike stikha [Articles on the poem linguistics]. Moscow: Slavic languages.
15. Gasparov M.L. 2001. Sintaksicheskaya struktura stikhotvornoy stroki [Syntactic structure of the verse line]. *Materialy mezhdunarodnoy konferentsii «Slavyanskiy stikh: Lingvisticheskaya i prikladnaya poetika»*. Moscow: Yazyki slavyanskoy kultury [The international conference "Slavonic verse: Linguistic and applied poetics], 130–137.
16. Gasparov M.L. 2012. Tochnyye metody analiza grammatiki v stikhe [Exact methods for grammar analysis in verses]. M.L. Gasparov. *Izbrannyye Trudy* [M.L. Gasparov. Selected works.], 4: 23–35.
17. Gasparov M.L. 2012. Pushkin i problemy poeticheskoy formy: yazyk i stikh [Pushkin and problems of poetic form: language and verse]. M.L. Gasparov. *Izbrannyye Trudy* [M.L. Gasparov. Selected works.], 4: 117–129.
18. Shou Dzh.T. 2001. Somnitel'nyye «è» v rifmakh poselitseyskoy poezii Pushkina [Doubtful "è" in the rhymes of Pushkin's post-literary poetry]. *Slavyanskiy stikh. Lingvisticheskaya i prikladnaya poetika* [Slavic verse. Linguistic and applied poetics]. Moscow: Slavic languages. 267–274.
19. Popescu I-I, Chech R, Best K-H., Altmann G. 2013. Descriptivity in Slovak lyrics // *Glottology*. V. 4 (1), P. 92–104.
20. Andreyev V.S. 2015. Kvantitativnoye opisaniye individual'no-avtorskoy metaforicheskoy sistemy G. Longfello [Quantitative description of the individual author's metaphorical system G. Longfello]. *Izvestiya smolenskogo gosudarstvennogo universiteta* [Bulletin of the Smolensk State University]. 3(31): 201–210.
21. Andreev S.N., Borisov V.V. 2015. Linguistic Analysis Based on Fuzzy Similarity models/ In "Sequences in Language and Text". Iss. 69. P. 7–34./ Edited by G.K. Mikros, J. Macutek. Berlin/Boston: De Gruyter Mouton.
22. Can F., Patton J.M. 2010. Change of word characteristics in 20th-Century Turkish literature: A statistical analysis // *Journal of quantitative linguistics*. Vol. 17. No. 3. P. 167–190.
23. Juola P. 2007. Becoming Jack London // *Journal of Quantitative Linguistics*. Vol. 14. No. 2. P. 145–147.
24. Andreyev V.S. 2012. Modelirovaniye individual'nogo stilya (na osnove lingvisticheskikh kharakteristik) [Individual style modeling (based on linguistic characteristics)]. Moscow: Flinta–Science.
25. Bayevskiy V.S., Semenova N.A. 2004. Evolyutsiya liricheskogo stikha kak osnova periodizatsii tvorcheskoy biografii poeta: klasternyy i korrelyatsionnyy analiz (Pushkin, Gumilev, Pasternak) [Evolution of the lyrical verse as the basis for the periodization of the poet's creative biography: cluster and correlation analysis (Pushkin, Gumilev, Pasternak)]. *Slavyanskiy stikh VII: Lingvistika i struktura stikha* [Slavic verse VII: Linguistics and the structure of the verse]. Edited by M.L. Gasparova, T.V. Skulachevoy. Moscow: Slavic languages. 421–436.
26. Krasnoperova M.A. 2000. Osnovy rekonstruktivnogo modelirovaniya stikhoslozheniya: Na materiale ritmiki russkogo stikha [Fundamentals of reconstructive modeling of versification: On the material of the rhythm of Russian verse]. St. Petersburg.: St. Petersburg State University.
27. Matyash S.A. 2006. Perenosy (enjambements) v lirike A.S. Pushkina i M.YU. Lermontova [Enjambements in the lyrics of A.S. Pushkin and M.Yu. Lermontov]. *Vestnik Orenburgskogo gosudarstvennogo universiteta* [Bulletin of the Orenburg State University]. 11:57–63.
28. Krishnapuram R., Keller J.M. A possibilistic approach to clustering // *IEEE Transactions on Fuzzy Systems*. 1993. Vol. 1. No. 2. P. 98–110.
29. Krishnapuram R., Keller J.M. The possibilistic C-Means Algorithm: Insights and Recommendations // *IEEE Transactions on Fuzzy Systems*. 1996. Vol. 4. No. 3. P. 385–393.
30. Borisov V.V., Andreyev S.N., Fedulov YA.A. 2014. Analiz slozhnykh lingvisticheskikh ob'yektov na osnove nechetkikh otsenochnykh modeley [Complex Linguistic Objects Analysis based on Fuzzy Estimating Models]. *Iskusstvennyy intellekt i prinyatiye resheniy* [Artificial intelligence and decision making]. 3:95–107.
31. Borisov V.V., Zernov M.M., Fedulov YA.A. 2013. Sposob nechetkogo mnogokriterial'nogo otsenivaniya s uchetom soglasovannosti parametrov otsenki [The method of fuzzy multi-criteria evaluation taking into account the consistency of the evaluation parameters]. *Informatsionnyy byulleten' Smolenskogo regional'nogo otdeleniya Akademii voyennykh nauk* [Bulletin of the Smolensk Regional Branch of the Academy of Military Sciences]. 28: 22–135.