В. С. Андреев

ДИНАМИКА СТИЛЯ Э. А. ПО
(на материале лирики)

Анализируются изменения лингвистических параметров в лирике Э. А. По при помощи многомерного дискриминантного анализа. Исследование позволяет выявить характеристики, которые дифференцируют различные периоды творчества поэта, определить степень сходства и различия между периодами, выделить тексты, являющиеся наиболее типичными репрезентантами каждого из этапов творчества поэта.

Ключевые слова: индивидуальный стиль, дискриминантный анализ, стихотворный текст, Эдгар Аллан По.

V. Andreev

E. A. POE'S STYLE IN DYNAMICS
(on the Material of Lyrics)

The changes of linguistic parameters in lyrics of E. A. Poe are studied. Multivariate discriminate analysis is applied. The characteristics differentiating various periods of the poet's life are described, they allow to identify the degree of similarity and difference between these periods and the texts which most typically represent every stage of the poet's creative life.

Keywords: Individual style, discriminant analysis, verse text, Edgar Allan Poe.

Сопоставление особенностей индивидуального стиля авторов художественных произведений является одним из наиболее интенсивно разрабатываемых вопросов в стилеметрии. Он рассматривается в исследованиях, относящихся к проблеме атрибуции текстов [8; 11; 19], к классификации текстов по гендеру, жанру [7; 16; 18], в рамках теории выдвижения [13; 17] и при решении ряда других задач.

Важной, хотя и значительно менее изученной областью, является динамический аспект идиостиля, степень изменчивости языковых характеристик в текстах, созданных одним и тем же автором в различное время.

Дж. Тамбуратзис и его коллеги среди 86 признаков, привлекаемых ими для дифференциации стиля сенаторов греческого парламента, в качестве основных, значительно влияющих на точность дискриминации, определили очередность произнесения речи во времени [20]. Существенные колебания в представленности ряда ритмических, синтаксических и других признаков в различных периодах творчества Пушкина, Гумилева и Пастернака были выявлены в работах В. С. Баевского [3]. В этих и других не очень многочисленных исследованиях, посвященных динамике идиостиля, получены результаты, свидетельствующие о возможности привлечения лингвистических параметров для дифференциации различных этапов творчества авторов [4; 12].

В нашей статье ставится задача определить, дифференцируют ли лингвистические характеристики текстов известного американского поэта-романтика XIX века Эдгара Аллана По различные периоды его творчества, и, если такая дифференциация возможна, выявить степень сходства и различия между периодами. Еще одна задача работы состояла в том, чтобы определить те тексты Э. По, которые являются наиболее типичными репрезентантами в каждом из периодов его творчества.

Выбор Э. По объясняется его ролью в американской и мировой литературе: он входит в число самых известных в мире американских поэтов и самых ярких авторов эпохи романтизма, его произведения переведены на многие языки мира и включены в большое число антологий.

Материалом исследования стали сопоставимые по объему лирические произведения По (не менее 8 и не более 70 строк), написанные одним и тем же метром — ямбом. Выбор текстов данного жанра объясняется тем, что именно в лирике в наиболее полной мере проявляется своеобразие творческой манеры поэта, а ямб является наиболее широко используемым американскими романтиками метром.

Тексты берутся по прижизненным сборникам лирических произведений Э. По. Публикации автором сборников стихов мы рассматриваем как важные вехи биографии автора, фиксирующие

особенности его творчества и стиля на определенный момент. Эти события неразрывно связаны с творческим процессом, являясь фактически его частью, а включенные в сборники тексты, отобранные самим автором, отражают, кроме всего прочего, его подсознательные языковые предпочтения на тот момент. Важность учета поэтических циклов, отражаемых в таких собраниях работ подчеркивается в целом ряде исследований [6; 10]. Всего Э. По издал четыре сборника лирики «Tamerlane and Other Poems» (1827 г.), «Al Aaraaf, Tamerlane and Minor Poems» (1829 г.), «Poems» (1831 г.), «The Raven and Other Poems» (1845 г.).

Для проведения сопоставительного анализа языковых характеристик в текстах Э. По мы выделяем три периода его творчества: «ранний», «средний» и «зрелый».

Первый, ранний, период жизни поэта относится ко времени 20-х годов, когда он делал свои первые шаги в литературе. Его ранние произведения, изданные небольшими тиражами, не произвели сколько-нибудь значительного впечатления на общественность. Конец первого периода и начало следующего совпадает со смертью приемной матери поэта и с повторной свадьбой отчима в 1830 году, что приводит к полному разрыву какой-либо связи Э. По с его отчимом и началу его самостоятельной жизни.

Второй период длится с 1830 по 1835–1836 годы. Его конец маркирован временем, когда Э. По становится редактором журнала «Southern Literary Messenger», получив постоянную связанную с литературой работу и возможность регулярно публиковать свои произведения. Последний период творчества Э. По (со второй половины 30-х годов и до конца жизни) отмечен значительным преобладанием работы над прозаическими произведениями и литературной критикой. Э. По публикуется в различных журналах и издает главным образом сборники прозаических произведений.

Исходя из указанной периодизации, четыре поэтических сборника Э. По распределяются по периодам следующим образом. К первому периоду относятся сборники «Tamerlane and Other Poems», «Al Aaraaf, Tamerlane and Minor Poems», ко второму периоду — сборник «Poems», а «The Raven and Other Poems» — к третьему.

Следует отметить, что Э. По иногда публиковал произведения повторно. В тех случаях, когда произведение было опубликовано им без изменений или с незначительными изменениями, которые не отражаются на вариативности используемых нами параметров (изменение запятой на тире и наоборот и т. д.), текст рассматривается один раз и относится к тому сборнику, в котором был опубликован впервые. В случае более значительных изменений, внесенных автором в повторно публикуемое произведение, мы рассматриваем отличающиеся редакции как разные тексты. Общий объем привлекаемых текстов составляет более 900 стихотворных строк (стихов).

К исследованию было привлечено 27 характеристик, включающих морфологические, синтаксические, макроструктурные признаки, а также признаки поэтического синтаксиса.

Морфологические характеристики (всего 10) определяются частеречной принадлежностью слов. Они основаны на подсчете количества существительных, глаголов, прилагательных, наречий, местоимений в двух важных позициях в стихе — в первой и последней сильных позициях.

Синтаксические признаки (12) описывают синтаксическую функцию слов, находящихся в указанных выше позициях.

В состав макроструктурных входят признаки «количество простых предложений в составе сложносочиненных

предложений», «количество придаточных в составе сложноподчиненных предложений».

Кроме того, нами используются параметры «стихотворного синтаксиса». В их число входят такие признаки, как «количество стихотворных переносов», «количество разрывов строки синтаксической паузой», а также количество стилистически маркированных (восклицательным, вопросительным знаком, многоточием) строк. Перенос (enjambement) берется в его традиционном значении — как перенесение части синтаксически целой фразы из одной строки стихотворного текста в другую: *I kneel, <... > / Amid thy shadows, and so drink within / My very soul thy grandeur, gloom, and glory!* (“The Coliseum”).

Разрыв стиха синтаксической паузой понимается как наличие в теле стиховой строки паузы, возникающей на границе разных предложений, простых предложений в составе сложного, вводных слов или вводных синтаксических конструкций, обращений, а также при наличии разделительных союзов.

Используемая здесь признаковая схема, учитывающая важные для анализа стихотворной речи параметры [9], была получена нами в результате отбора признаков из большого числа характеристик в результате исследований по классификации стихотворных текстов [1; 2].

Поскольку произведения имеют разный объем, количественные данные по каждому тексту были нормированы путем деления количества зафиксированных случаев реализации каждого из 27 признаков в стихотворении на количество строк этого стихотворения.

Для решения поставленной задачи в этом исследовании, как и в ряде наших других работ, посвященных классификации и анализу параметров идиостиля, мы применяем многомерный дискриминантный анализ [5]. Этот метод в настоящее время начинает все чаще успешно применяться в лингвистике для атрибуции и автоматической классификации текстов [14; 15]. Его преимуществом является то, что он позволяет сопоставить группы объектов, принадлежащих различным классам, одновременно по значительному числу признаков с целью выявления среди них таких комплексов характеристик, которые дифференцируют сопоставляемые классы. В нашем случае объектами являются тексты лирических стихотворений Э. По, опубликованные им в сборниках, а классами — три периода творчества Э. По.

Дискриминантный анализ использовался нами для решения следующих конкретных задач: а) выявление степени сходства или различий между периодами творчества Э. По по указанным выше лингвистическим параметрам текстов; б) выделение произведений, отражающих в наиболее полной форме лингвистические маркеры каждого периода.

Творческий путь Э. По был достаточно короток, что могло бы в принципе препятствовать значительным изменениям в идиостиле автора и соответственно создать дополнительные трудности в поиске каких-либо возможных изменений. Однако в результате проведения дискриминантного анализа было установлено, что три выделенных периода имеют существенные различия. Из общего списка привлекаемых к анализу параметров 14 оказались обладающими дискриминантной силой относительно периодов творчества поэта. Указанные признаки составляют дискриминантную «модель», т. е. различают (дифференцируют) три периода творчества Э. По. Наиболее значимыми для дискриминации оказались признаки, отражающие количество стилистически маркированных строк, количество наречий, существительных, глаголов, определений в конце строки, количество глаголов, местоимений, прила-

гательных в начале строки, количество придаточных предложений в составе сложноподчиненных предложений.

Представляется возможным определить, какова точность дифференциации, основанной на модели. Для этого можно осуществить проверку, проведя автоматическое разбиение текстов на группы на основании вошедших в модель признаков. Суть этой проверки будет состоять в том, что программа автоматически разобьет тексты Э. По на три периода, «не зная», к какому реально периоду относится тот или иной текст, а основываясь исключительно на признаках выявленной дискриминантной модели. После того как группировка будет осуществлена, ее можно сравнить с тем, как в действительности произведения распределены по периодам творчества поэта. Чем больше произведений в результате автоматической группировки окажется в «правильных» классах, т. е. будет отнесено к тем периодам, к которым они относятся в реальности, тем более точной является выявленная дискриминантная модель.

В табл. 1 приводятся данные о группировке текстов по описанной методике. По горизонтали указаны классы текстов согласно естественной группировке (принадлежности данному периоду творчества). По вертикали классы даны так, как они образуются в результате группировки текстов в пространстве признаков модели.

Полученные результаты оказались очень высокими с точки зрения точности дискриминантной модели, достигнув максимального показателя. При случайном распределении текстов для трех классов, т. е. в случае, если бы модель была совершенно не эффективна для различения периодов, можно было бы ожидать около 33% правильного отнесения [5]. Таким образом, полученные нами результаты являются весьма убедительным доказательством эффективности модели. Они свидетельствуют о том, что, во-первых, существует значимое различие между языковыми параметрами текстов Э. По на различных этапах его творчества и, во-вторых, что отобранные признаки позволяют с очень высокой точностью эти различия выявить.

Установив факт дифференциации этапов творчества Э. По в рамках используемой лингвистической модели, можно перейти к их сопоставлению. В табл. 2 отражена степень различия текстов, относящихся к трем периодам.

Показателем сходства/различия периодов является расстояние между «центроидами» групп текстов каждого из трех периодов. Тексты рассматриваются как точки в многомерном пространстве, образуемом привлекаемыми к исследованию признаками. Под центроидом периода понимается точка в пространстве с координатами, являющимися средними значениями переменных (признаков) всех текстов в данной группе. Чем больше расстояние между центроидами, тем больше различие между соответствующими периодами.

Таблица 1

Распределение текстов трех периодов в пространстве признаков модели

Исходные классы текстов	Процент правильного распределения	Распределение текстов согласно признаковой модели в процентах		
		Период 1	Период 2	Период 3
Период 1	100	100	0	0
Период 2	100	0	100	0
Период 3	100	0	0	100
Всего	100	100	100	100

Таблица 2

Расстояние между центроидами различных периодов творчества Э. По

	Период 1	Период 2	Период 3
Период 1	0,00	21,25	28,97
Период 2	21,25	0,00	16,04
Период 3	28,97	16,04	0,00

Расстояния между центроидами определялось при помощи меры Махаланобиса (квадрат расстояния Махаланобиса). Полученные значения округлены до сотых долей.

Исходя из значений табл. 2, можно сделать вывод о противопоставлении первого периода двум другим, причем максимальное различие имеет место между первым и последним периодами. С другой стороны, второй период удален от третьего в значительно меньшей степени. Можно заключить, что основные изменения в представленности признаков дискриминантной модели имели место в конце 20-х и начале 30-х годов XIX века, однако затем в течение всех последующих лет эти лингвистические параметры были более стабильными в лирике Э. По.

Выделение центроидов периодов позволяет не только определить различия между периодами, но и оценить тексты с точки зрения их принадлежности к ядру или периферии соответствующего этапа творчества поэта. К ядру относятся тексты, наиболее близкие к центроиду, то есть обладающие характеристиками, близкими к усредненным. Эти произведения могут рассматриваться как отражающие «архетип» данного периода. Периферия, напротив, сформирована удаленными от центроида текстами, признаки которых значительно отличаются от средних.

Рассмотрим произведения, в наибольшей степени отражающие архетипы своих классов.

Для первого периода этими текстами оказались следующие произведения: «A Dream» (сборник «Tamerlane and Other Poems») и «The Lake — To —» (сборник «Al Aaraaf, Tamerlane and Minor poems»). Во втором периоде творчества ближе всего к центроиду располагается произведение «A Paean». Наиболее близкими к центроиду третьего периода являются тексты «Israfel» и «The Sleeper».

В текстах, составляющих ядро первого периода, присутствуют характерные для периода признаки: достаточно высокое количество придаточных в сложноподчиненных предложениях, глаголов в финальной части строки и определений в конце строки.

Прототипический текст второго периода демонстрирует большое количество существительных в финальной позиции в строке (*They loved her for her wealth — / And they hated her for her pride —*) и глаголов в первой сильной позиции (*I strike — the murmur sent / Through the grey chambers to my song, / Shall be the accompaniment*).

Тексты, составляющие ядро третьего периода, отличаются значительным количеством строк с синтаксически маркированным концом: *The lady sleeps! Oh, may her sleep, / Which is enduring, so be deep! / Heaven have her in its sacred keep!* (“The Sleeper”).

Подводя итоги, можно сделать следующие основные выводы.

Результаты проведенного анализа показывают, что, несмотря на относительно короткий период творческой активности Э. По, на базе синтаксических и морфологических признаков его лирических произведений удается дифференцировать различные этапы творчества поэта с высокой точностью.

Используемая нами методика позволила определить, что наибольшие изменения имели место в идиостиле Э. По между первым и вторым периодами творчества, в конце 20-х годов XIX века.

Привлечение дискриминантного анализа позволило также установить, какие тексты репрезентируют в наибольшей степени каждый из этапов.

СПИСОК ЛИТЕРАТУРЫ

1. *Андреев В. С.* Классификация стихотворных текстов методом дискриминантного анализа (на материале лирики американских поэтов-романтиков) // Вестник Минского государственного лингвистического университета. Серия 1. Филология. — Минск: МГЛУ, 2002. №10. С. 141–146.

2. *Андреев В. С.* Динамика идиостиля Дж. Г. Уитьера // Вестник Тамбовского университета. Серия «Гуманитарные науки». Тамбов, 2007. Вып. 12(56). С. 266–270.

3. *Баевский В. С.* Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы. — М.: Языки славянской культуры, 2001. — 336 с.

4. *Карелова О. В.* Язык и стиль поэзии Роберта Лоуэлла: Автореф. дис. … канд. филол. наук. — СПб., 2007. — 18 с.

5. *Клекка У. Р.* Дискриминантный анализ // Факторный, дискриминантный и кластерный анализ. — М.: Финансы и статистика, 1989. С. 178–138.

6. *Ляпина Л. Е.* Циклизация в русской литературе XIX века. — СПб., 1999. — 279 с.

7. *Мартыненко Г. Я.* Основы стилеметрии. — Л.: Изд-во Ленингр. ун-та, 1988. — 174 с.

8. *Марусенко М. А.* Атрибуция анонимных и псевдонимных литературных произведений методами распознавания образов. — Л.: Изд-во Ленинградского университета, 1990. — 168 с.

9. *Скулачева Т. В.* Ритм и грамматика в стихе // Славянский стих. Лингвистическая и прикладная поэтика: Лингвистическая и прикладная поэтика: Материалы международной конференции 23–27 июня 1998 г. — М.: Языки славянской культуры, 2001. С. 121–129.

10. *Фоменко И. В.* Лирический цикл. Становление жанра, поэтика. — Тверь, 1992. — 124 с.

11. *Baayen R. H., Van Halteren H., Tweedie F.* Outside the cave of shadows: Using syntactic annotation to enhance authorship attribution // Literary and Linguistic Computing, 1996. Vol. 11. P. 121–131.

12. *Can F., Patton J. M.* Change of Writing Style with Time // In: Computers and the Humanities, 2004. Vol. 38. P. 61–82.

13. *Hakemulder J.* Foregrounding and its Effects on Readers' Perception // Discourse Processes, 2004. Vol. 38. P. 193–218.

14. *Holmes D. I., Forsyth D.* The Federalist Revisited: New Directions in Authorship Attribution // Literary and Linguistic Computing, 1995. Vol. 10. P. 111.

15. *Karlgen J., Cutting D.* Recognizing text genres with simple metrics using discriminant analysis // Proceedings of COLING 94, Kyoto, 1994. P. 1071–1075.

16. *Koppel M., Argamon S., Shimoni A. R.* Automatically Categorizing Written Texts by Author Gender // Literary & Linguistic Computing, 2002. Vol. 17. P. 401–412.

17. *Miall D. S., Kuiken D.* Shifting Perspectives: Readers' Feelings and Literary Response / W. van Peer and S. Chatman (eds) New Perspectives on Narrative Perspective: Albany, NY: SUNY Press, 2001. P. 289–301.

18. *Murata M.* Identify a Text's Genre by Multivariate Analysis — Using Selected Conjunctive Words and Particle-phrases // Proceedings of the Institute of Statistical Mathematics, 2000. Vol. 48. P. 311–326.

19. *Stamatatos E., Fakatakis N., Kokkinakis G.* Computer-Based Authorship Attribution Without Lexical Measures // Computers and the Humanities, 2001. Vol. 35. P. 193–214.

20. *Tambouratzis G., Markantonatou S., Hairetakis N., Vassiliou M., Carayannis G., Tambouratzis D.* Discriminating the Registers and Styles in the Modern Greek Language — Part 2: Extending the Feature Vector to Optimize Author Discrimination // Literary and Linguistic Computing, 2004. Vol. 19. № 2. P. 221–242.