Moscow, 1981; *Pavi P.* Slovar teatra. Moscow, 1991; *Ukolova L.E.* Drama kak literaturnyj rod: svoïstva i otnoshenija. Dnepropetrovsk, 1999; *Halizev V.E.* Drama kak rod literatury (poetika, genezis, funkcionirovanie). Moscow, 1986.

[2] *Гачев Г.Д.* Указ. соч. С. 197.

Gachev G.D. Op. cit. P. 197.

[3] *Grünewald H.* Generic Ambiguity in Daniil Kharms's "Elizaveta Bam" // New Zealand Slavonic Journal. 2001. P. 87–99; *Roberts G.H.J.* Of Words and Worlds: Language Games in Elizaveta Bam by Daniil Kharms // The Slavonic and East European Review. 1994. Vol. 72. № 1 (Jan.). P. 38–59; *Жаккар Ж.-Ф.* Даниил Хармс и конец русского авангарда. М., 1995; *Красильникова Е.Г.* Русская авангардистская драма: человек отчужденный // Русская литература. 1998. № 3. С. 57–67; *Токарев Д.В.* Курс на худшее: Абсурд как категория текста у Д. Хармса и С. Беккета. М., 2002; *Шенкман Я.С.* Логика абсурда (Хармс: отечественный текст и мировой контекст) // Вопросы литературы. 1998. № 4. С. 54–81.

Grünewald H. Generic Ambiguity in Daniil Kharms's "Elizaveta Bam" // New Zealand Slavonic Journal. 2001. P. 87–99; *Roberts G.H.J.* Of Words and Worlds: Language Games in Elizaveta Bam by Daniil Kharms // The Slavonic and East European Review. 1994. Vol. 72. № 1 (Jan.). P. 38–59; *Zhakkar Zh.-F.* Daniil Harms i konec russkogo avangarda. M., 1995; *Krasilnikova E.G.* Russkaja avangardistskaja drama: chelovek otchuzhdennyï // Russkaja literatura. 1998. № 3. P. 57–67; *Tokarev D.V.* Kurs na hudshee: Absurd kak kategorija teksta u D. Harmsa i S. Bekketa. Moscow, 2002; *Shenkman Ja.S.* Logika absurda (Harms: otechestvennyj tekst i mirovoj kontekst) // Voprosy literatury. 1998. № 4. P. 54–81.

[4] Манифест ОБЭРИУ // Даниил Хармс. Творчество Даниила Хармса. URL: http://xapмc.gorodok.net/documents/1423/default.htm (дата обращения: 07.04.2013).

Manifest OBERIU // Daniil Harms. Tvorchestvo Daniila Harmsa. URL: http://xapмc.gorodok.net/documents/1423/default.htm (accessed: 07.04.2013).

[5] *Фарино Е.* Введение в литературоведение. СПб., 2004.

Farino E. Vvedenie v literaturovedenie. St. Petersburg, 2004.

[6] *Хармс Д.И.* Полет в небеса: Стихи. Проза. Драмы. Письма. Л., 1988. С. 520.

Harms D.I. Polet v nebesa: Stihi. Proza. Dramy. Pisma. Leningrad, 1988. P. 520.

[7] Манифест ОБЭРИУ.

Manifest OBERIU.

[8] Там же.

Ibidem.

[9] Там же.

Ibidem.

[10] Там же.

Ibidem.

[11] *Хармс Д.И.* Полет в небеса. С. 482.

Harms D.I. Polet v nebesa. P. 482.

[12] *Хармс Д.И.* Горло бредит бритвою: случаи, рассказы, дневниковые записи // Глагол. Литературно-художественный журнал. 1991. № 4. С. 205.

Harms D.I. Gorlo bredit britvoju: sluchai, rasskazy, dnevnikovye zapisi // Glagol. Literaturno-hudozhestvennyj zhurnal. 1991. № 4. P. 205.

[13] *Тамарченко Н.Д., Тюпа В.И., Бройтман С.Н.* Сюжет в эпоху синкретизма // Теория литературы: В 2 т. / Под ред. Н.Д. Тамарченко. М., 2004. Т. 2. С. 51–72.

Tamarchenko N.D., Tjupa V.I., Brojtman S.N. Sjuzhet v epohu sinkretizma // Teorija literatury: V 2 t. / Pod red. N.D. Tamarchenko. Moscow, 2004. T. 2. P. 51–72.

[14] *Токарев Д.В.* Указ. соч. С. 157, 190–191.

Tokarev D.V. Op. cit. P. 157, 190–191.

[15] *Хармс Д.И.* Полет в небеса. С. 483.

Harms D.I. Polet v nebesa. P. 483.

ПРОЧТЕНИЯ

Б.В. Орехов, И.В. Пешков (Москва)

НА ПОДСТУПАХ К ГЕНОМУ СТИЛЯ ШЕКСПИРА

В работе предлагается оригинальная методика представления и анализа стиля. Стиль определяется как совокупность всех повторных коллокаций произведений автора. Методика применяется к корпусу пьес Шекспира. Из текстов, включенных в Первое Фолио, было получено два словника коллокаций с любой, вплоть до минимальной (присутствие хотя бы в двух пьесах корпуса), и максимальной (присутствие во всех 36 пьесах) повторностью единиц. Первый словник составил более 130000, второй 218 единиц. Сравнение с этими словниками отдельных произведений Шекспира, современников Шекспира и других авторов, писавших на английском языке, дало возможность авторам статьи сделать некоторые выводы об атрибуции отдельных художественных произведений и частей этих произведений, в частности подтвердить и уточнить феномены соавторства при создании определенных шекспировских пьес.

Ключевые слова: стиль; коллокация; Шекспир; авторство; атрибуция; Первое Фолио; Эдвард де Вер; драматурги-елизаветинцы; шекспировский вопрос.

Казалось бы, при том внимании, которое традиционно уделяется Шекспиру, за прошедшие четыре сотни лет филологи и историки обязаны были до чего-то докопаться, а основные проблемы, по крайней мере, те, что вообще имеют решение, должны были быть разрешены (скорее всего, так должен полагать обычный читатель или зритель). Ничуть не бывало. Шекспир со всеми его загадками оказался человеку не по зубам. Осознав это, человек призвал себе на помощь компьютер. В 1990 г. был сделан показательный обзор[1] уже трех существовавших к тому времени электронных изданий произведений Шекспира и программ, позволявших с этими изданиями работать, решая, впрочем, довольно примитивные по нынешним меркам задачи. Понятно, что с тех пор число приложений компьютерных технологий к исследованиям шекспировских текстов только росло – вместе с доступностью техники и ее вычислительными мощностями. При этом сам по себе компьютерный анализ (в том числе и анализ текста) качественно не отличается от того, который выполняется людьми. Да и методику для такого анализа тоже разрабатывает человек. Что-то находится, систематизируется, подсчитывается. Отличие только в том, что компьютер может обработать за единицу времени гораздо больше материала и тут уже есть шанс, что количество перейдет в качество: результаты подсчетов будут весомее, а выводы, сделанные на их основе (уже снова человеком), – устойчивее.

Разумеется, более всего компьютерных лингвистов интересовала проблема атрибуции текстов и, соответственно, проблема стиля[2]. Какое бы

значение ни вкладывать в это слово, ясно, что к проблеме авторства стиль имеет самое прямое отношение. И тут уже все средства хороши: и нейронные сети, и анализ модальности[3], и многие другие. Выводы основываются на строгой статистике, только подсчеты ведутся каждый раз по разным параметрам.

Кроме того, с Шекспиром сыграла компьютерную шутку его репутация основателя английской литературы. Постоянно находящиеся в центре внимания, известные всем и часто цитируемые, шекспировские тексты чаще других и становятся площадкой для испытания новых методик компьютерной лингвистики, будь то извлечение из текста его фрактальной геометрии[4] или контент-анализ с целью описания эстетического эффекта[5]. Так что испытывать новый способ извлечения из текста некоторых его единиц и характеристик в первую очередь именно к Шекспиру – дело уже привычное.

И несмотря на все это, несмотря на столь давнюю, с точки зрения компьютерной истории традицию объединения в филологических штудиях Шекспира и компьютера, в нашей работе сложности возникли еще до того, как мы начали их осознавать. Прежде всего, оказалось трудно сформулировать саму задачу наших исследований. Формулировка «Определить стиль Шекспира», а именно к этому мы интуитивно стремились, сразу оказалась размытой, причем во всех своих терминах. Во-первых, нет достаточной ясности, что такое стиль. Это отдельная теоретическая проблема. Во-вторых, нет никакой ясности, кто такой Шекспир: это человек, бренд или псевдоним? В-третьих, наконец, при таких вводных становится неясным, что вообще значит «определить»…

Но поскольку решать задачу из трех неизвестных при отсутствии известных невозможно, то пришлось кое-что признать известным. Вот мы и решили, что исходя из своего исследовательского, читательского и просто житейского опыта кое-что уже знаем о стиле. Мы принимаем за лемму утверждение:

Стиль – это совокупность типичных сочетаний слов
(исходная лемма А).

Уже чуть легче. Осталось найти сочетания слов у Шекспира и определить степень их типичности. Но тут же встает во весь рост вторая проблема: кто такой Шекспир? Теперь мы уже не можем назначать себе вторую лемму и исходить из того, что Шекспир это тот-то и тот-то, потому что лемма должна быть одна. И с одной-то не так все просто, а две леммы на старте рассуждений – это уже ни в какие ворота логики не войдет, поэтому от вопроса *кто такой Шекспир* в качестве исходного пункта нужно постараться просто уйти. В принципе это для шекспироведов традиционный уход: основная филологическая работа сосредоточена почти исключительно на произведениях Шекспира. То есть и мы начнем в русле исследовательской традиции.

Таким образом, несмотря на знаменитое определение «стиль это человек», работать мы будем не с человеком, а с текстом. Вернее, с корпусом текстов. Кстати, это практически единственное, что оставил после себя Шекспир. Корпус текстов. С этим корпусом тоже не все однозначно, но кое-что все-таки общепризнанно. Самым общепризнанным фактом в шекспироведении является то, что 36 пьес, изданных в 1623 г. в Первом Фолио, написал Шекспир. Так что мы просто считаем исходным текстовым материалом для определения стиля Шекспира пьесы Первого Фолио. Не будем углубляться в текстологию каждой пьесы (это отдельный вопрос), просто возьмем все тексты в их современной орфографии. (В качестве исходной точки нами было взято свободно распространяемое электронное издание пьес Шекспира в XML, подготовленное Джоном Босаком в 1999 г. на основе оцифрованных текстов Moby Lexical Tools[6]. Из входящих в собрание Босака пьес нами удален «Перикл», отсутствующий в Первом Фолио). Для предварительного, чернового анализа стиля этого достаточно и, более того, на этом нужно остановиться, помня об условности перевода в современную орфографию, но предполагая, что такой перевод произведен для всех анализируемых дальше текстов по одинаковому алгоритму. Для первой, черновой попытки стилистического анализа с этим предположением можно согласиться.

Итак, пока для нас нет самого Шекспира, но есть условно-полный текст Шекспира, из которого уже можно вычленять стиль, то есть определять сочетания слов. За слово в тексте примем последовательное сочетание букв между двумя небуквами (считая буквами также апостроф и дефис, а небуквами – любой другой типографский знак или пробел), то есть подходим к определению слова формально-графически. Для верности возьмем все сочетания (то есть рядоположения) от двух до шести слов и выберем те из них, которые появлялись хотя бы в двух пьесах Шекспира (этим и определяется типичность сочетаний). Вручную на эту операцию, вероятно, ушло бы полжизни, но компьютер делает все довольно быстро. В работу берутся только реплики персонажей. XML-разметка позволяет легко отсекать ремарки (скорее всего, в большинстве случаев не аутентично шекспировские), имена персонажей и прочие позднейшие пометы. Входящими в одно сочетание считаются даже слова, разделенные как границей строки, так и границей реплик разных персонажей. Значимой признается только граница акта. Таким образом, слово, которым оканчивается, например, I акт и слово, с которого начинается II акт, не будут входить в одно сочетание. (Хотя в принципе и от этой границы можно было бы отказаться.)

В результате мы получаем словник примерно из 133 тысяч *коллокаций*, так определяя сочетания слов, чтобы не путать с известным понятием

из синтаксиса (*словосочетание*). Этот результат назовем максимальным или Большим словником Шекспира (Бсл). На глаз для примерно 15 тысяч слов в языке автора[7] 130 тысяч типичных коллокаций вполне нормально.

Теперь выберем самые типичные шекспировские коллокации: те, что имеются во всех 36 пьесах. Словник сузился до 218 единиц. Да, их мало, самых типичных, но они есть! Вот 10 самых-самых типичных сочетаний слов у Шекспира (правый крайний столбик показывает, сколько раз коллокация встретилась в корпусе):

1. I am		1814
2. I have		1587
3. in the		1557
4. I will		1550
5. to the		1476
6. of the		1355
7. it is		1069
8. to be		939
9. that I		911
10. I do		810

Конечно, даже начинающему изучать английский язык сразу понятно, что перед нами одни из самых частотных сочетаний слов в языке вообще, и на первый взгляд кажется, что оригинальность стиля Великого барда с помощью этого, назовем так, Малого словника (Мсл) Шекспира определить будет невозможно. По простому наличию этих коллокаций в любом тексте, разумеется, ничего не скажешь о стиле, зато по частотности разных коллокаций этого словника кое-что о стиле сказать будет можно. Для того, чтобы определить, насколько синтаксис текста (а малый словник явно отвечает за синтаксис, ибо состоит по преимуществу из служебных слов типа артиклей, предлогов, личных местоимений и глагольных связок!) приближается к шекспировскому, мы попробовали оценивать сумму мест по Мсл (где единицы последовательно расположены от максимальной частоты вхождений к минимальной) первых ста коллокаций анализируемого текста. Например, так будет выглядеть начало списка для «Алисы в Стране Чудес» Л. Кэрролла:

N	N(Мсл)	collocation	quant	quant(Мсл)
1	(6)	of the	125	1355
2	(47)	in a	97	436
3	(3)	in the	79	1557
4	(12)	and the	77	707
5	(5)	to the	69	1476
	(= 73)			
6	(79)	at the	60	329
7	(8)	to be	48	939
8	(55)	on the	34	413
9	(26)	with the	33	532
10	(133)	and then	29	235
	(= 374)			
...				

Здесь N – позиция коллокации в частотном списке у Кэрролла, N(Мсл) – место той же коллокации в Мсл, collocation – сама коллокация, quant – абсолютная встречаемость в «Алисе», quant(Мсл) – частота употреблений у Шекспира. В скобках со знаком «=» после пятой и десятой позиции – сумма мест по малому словнику.

В кратком виде, продолженном до 100-й позиции, это можно представить так:

Nposit	positSum
5	73
10	374
15	790
20	1313
25	1795
30	2385
40	3412
50	4592
75	7056
100	9284

Здесь Nposit – это позиция у Кэрролла, количество коллокаций в частотной для этого текста последовательности (от большей частоты употребления к меньшей), positSum – сумма мест, занимаемых этими коллокациями в Мсл.

А так эти же списки будут выглядеть для «Ромео и Джульетты»:

N	N(Мсл)	collocation	quant	quant(Мсл)
1	(4)	I will	60	1550
2	(1)	I am	53	1814
3	(3)	in the	41	1557
4	(2)	I have	37	1587
5	(7)	it is	34	1069

```
     (= 17)
6    (6)   of the                33      1355
7    (5)   to the                31      1476
8    (15)  is the                30      676
9    (31)  thou art              27      508
10   (126) is my                 23      250
     (= 200)
...
```

Nposit	positSum
5	17
10	200
15	327
20	560
25	717
30	938
40	1650
50	2176
75	3883
100	6756

Чем меньше сумма мест у определенного числа коллокаций, скажем десяти, тем ближе «синтаксис» текста к шекспировскому, поскольку последовательность наиболее частотных слов больше соответствует Мсл.

Собственно говоря, это весь инструментарий. Большой шекспировский словник, состоящий из 133 тысяч единиц (Бсл) и малый шекспировский словник, состоящий из 218 единиц (Мсл), – это то, с чем уже можно сравнивать реальные тексты, прежде всего на предмет наличия в них единиц этого словника.

Мы сравнивали со словниками три группы текстов:

1) каждую из 36 пьес Шекспира (корпус по 1 Фолио);

2) произведения современников, ближайших предшественников (пока собственно только «Кентерберийские рассказы» Дж. Чосера) и ближайших литературных потомков («Потерянный рай» Мильтона);

3) более поздние тексты на английском языке самых разных жанров (поэзия Байрона, Шелли, проза Стерна, Филдинга, Дефо, Свифта, Диккенса, Теккерея, Кэрролла, Стивенсона, Лондона, Элиот, Мелвилла, Марка Твена, Агаты Кристи, Герберта Уэллса, Андре Нортон, а также «История затмений» Чамберса, «Золотая ветвь» Фрэзера, «Очерки молочной бактериологии» Рассела).

Для каждого произведения мы посчитали количество присутствующих шекспировских коллокаций по Бсл и Мсл, а также их плотность, то есть частоту употребления на единицу текста. Плотность присутствия единиц словника (P) получалась простым делением числа *разных* коллокаций на количество слов в тексте, а плотность общего количества употреблений (P_{use}) этих единиц (каждую коллокацию в тексте можно было употребить не один раз) получалась делением числа всех шекспировских коллокаций, найденных в тексте, на количество слов.

В первой группе текстов мы получили типичные шекспировские показатели, последовательно обработав все пьесы. Так, например, выглядит картина по «Антонию и Клеопатре» (23684 слова):

```
шекспировских коллокаций  из Бсл: 14980
14980 : 23684  = 0.632 (P)
21659 : 23684 = 0.914 (Puse)
шекспировских коллокаций  из Мсл: 218
218 : 23684  = 0.009 (P)
1975 : 23684 = 0.083 (Puse)
```

Первые две операции деления относятся к Бсл, а последние две – к Мсл. Наиболее стилеразличающие показатели – это P по Бсл и P_{use} по Мсл. Особенно несущественным пока представляется показатель P по Мсл, поскольку при достаточных объемах текста он зависит почти исключительно от общего количества анализируемых слов.

А вот как для «Антония и Клеопатры» выглядит таблица суммы мест по Мсл:

Nposit	positSum
5	19
10	107
15	231
20	415
25	608
30	794
40	1226
50	2083
75	3705
100	6169

Так представлены результаты анализа. По первой, шекспировской, группе текстов выведены средние показатели: сумма P была разделена на 36, по числу пьес в Первом Фолио. (Далее в целях апробирования методики мы применяли ее к отрезкам текста в 2000, 5000 и 30000 (или сколько есть в произведении) слов. В таких случаях среднее выводилось делением суммы показателей на число отрезков).

для Бсл –

P = 0.**672**

P_{use} = 0.996

для Мсл –

P = 0.010

P_{use} = 0.**095**

Nposit	positSum
5.	**26**
10.	**109**
15.	227
20.	429
25.	622
30.	872
40.	1495
50.	**2151**
75.	4198
100.	6758

Все показатели текстов второй и третьей группы отличаются от шекспировских. То есть можно надеяться, что в результате описанных подсчетов получаются не случайные цифры, а системно отражающий внутренние характеристики текста набор численно выражаемых параметров, по которым можно сравнивать и различать тексты между собой.

Вторая группа распадается на тексты, которые существенно ближе к шекспировским по этим показателям, и все остальные. Возьмем сначала пример из «остальных»: первые 30 000 слов из «Кентерберийских рассказов»:

```
шекспировских коллокаций из Бсл: 7745
7745 : 30000  = 0.258
15756 : 30000 = 0.525
шекспировских коллокаций из Мсл: 191
191 : 30000  = 0.006
2396 : 30000 = 0.079
```

Nposit	positSum
5	81
10	333
15	528
20	769
25	1162
30	1458

40	2273
50	3071
75	6315
100	8862

И словарь (Бсл), и синтаксис (Мсл) существенно отличаются от шекспировских показателей.

Следующий пример уже из группы более близких по стилевым показателям к Шекспиру произведений (Бен Джонсон «Cynthia's Revels»):

```
шекспировских коллокаций из Бсл: 9951
9951 : 30000  = 0.359
17800 : 30000 = 0.648
шекспировских коллокаций из Мсл: 214
214 : 30000  = 0.007
2719 : 30000 = 0.094
```

Nposit	positSum
5	20
10	101
15	255
20	360
25	573
30	753
40	1400
50	1993
75	4158
100	6741

Коэффициент плотности по Бсл для этого произведения лишь чуть-чуть повыше, чем у Чосера, зато плотность употребления Мсл (0.094) почти идентична среднешекспировскому показателю (0.095), да и наиболее частотные коллокации по Мсл – тоже вполне шекспировские (*of the*, *in the*, *to the*, *of his*, *to be*), что предположительно объясняется сильным влиянием стиля Шекспира на Джонсона. На сознательном уровне Джонсон, конечно, этого не хотел и от прямого (лексического) подражания стилю Шекспира уходил, однако на уровне синтаксиса не избежал влияния. Так можно попытаться проинтерпретировать эти данные.

А вот анализ одной из пьес Кристофера Марло с наиболее шекспировскими (по сравнению с другими произведениями Марло) показателями («Мальтийский Еврей»):

```
шекспировских коллокаций из Бсл: 8834
8834 : 23312  = 0.426
14455 : 23312 = 0.669
шекспировских коллокаций из Мсл: 216
216 : 23312  = 0.010
1985 : 23312 = 0.086
```

Nposit	positSum
5	22
10	82
15	180
20	533
25	742
30	1004
40	1595
50	2500
75	5123
100	7773

Главный коэффициент, конечно, маловат, однако синтаксис до 50 коллокаций прямо-таки среднешекспировский!

Далее. Возьмем одну из многих пьес, изданных анонимно. «Edmund Ironside»:

```
шекспировских коллокаций из Бсл: 7101
7101 : 15599  = 0.461
10308 : 15599 = 0.669
шекспировских коллокаций из Мсл: 204
204 : 15599  = 0.013
1205 : 15599 = 0.078
```

Nposit	positSum
5	27
10	138
15	326
20	617
25	893
30	1245
40	2005
50	2747
75	4946
100	7181

По главному показателю (0.461) эта пьеса существенно ближе к Шекспиру, чем пьеса Джонсона, и несколько ближе, чем пьеса Марло. Не случайно это произведение постоянно фигурирует в числе претендентов на шекспировское авторство. Очень шекспировская сумма мест первого десятка коллокаций. Пьеса нуждается в более пристальном анализе.

Аналогичный случай с еще лучшими показателями по Мсл. («Damon and Pithias»):

```
шекспировских коллокаций из Бсл: 7565
7565 : 18546  = 0.407
12853 : 18546 = 0.693
шекспировских коллокаций из Мсл: 206
206 : 18546  = 0.011
1761 : 18546 = 0.094
```

Nposit	positSum
5	20
10	160
15	351
20	545
25	845
30	1108
40	1762
50	2477
75	4753
100	7641

Теперь из серии так называемых апокрифов, пьес, еще в шекспировское время или чуть-чуть позднее изданных под авторством Шекспира, которое потом было оспорено.

«Arden Of Feversham»:

```
шекспировских коллокаций из Бсл: 10730
10730 : 24686  = 0.434
17587 : 24686 = 0.712
шекспировских коллокаций из Мсл: 216
216 : 24686  = 0.008
2216 : 24686 = 0.089
```

Nposit	positSum
5	*20*
10	*137*
15	*286*
20	*429*
25	*586*
30	*849*
40	*1512*
50	*2096*
75	*4186*
100	*6877*

Достаточно высокий показатель по Бсл подкрепляется почти стопроцентным попаданием по Мсл. Другие апокрифы также демонстрируют бóльшую близость к шекспировским параметрам, чем другие тексты. Один из апокрифов уже давно признан шекспировским («Перикл»), хотя его показатели не намного лучше, чем у других апокрифов и даже немногим более шекспировские, чем у «Мальтийского Еврея» Марло:

```
шекспировских коллокаций из Бсл: 8502
8502 : 18365  = 0.462
12481 : 18365 = 0.679
шекспировских коллокаций из Мсл: 212
212 : 18365  = 0.011
1422 : 18365 = 0.077
```

Nposit	positSum
5	*19*
10	*103*
15	*263*
20	527
25	800
30	*922*
40	1563
50	*2380*
75	*4184*
100	*6746*

А вот тексты, заведомо считающиеся шекспировскими, на этих поэмах впервые напечатано имя «Шекспир» («Венера и Адонис» и «Обесчещенная Лукреция» вместе, потому что так они дают более релевантный для подсчетов объем):


```
шекспировских коллокаций из Бсл: 8363
8363 : 25213  = 0.331
12949 : 25213 = 0.513
шекспировских коллокаций из Мсл: 194
194 : 25213  = 0.007
1346 : 25213 = 0.053
```

Nposit	positSum
5	112
10	390
15	655
20	910
25	1577
30	1984
40	2930
50	3616
75	6391
100	9265

Ничего показательно шекспировского, в сравнении с апокрифами например, в них не обнаружилось. Волей-неволей остается задуматься либо над тем, «из чего состоит Шекспир»: это группа авторов под единым брендом или псевдоним (опять-таки одного человека или, скажем, учителя с учениками), либо над тем, как влияют на стиль жанровые особенности. Хотя не исключено, что и над тем, и над другим вместе.

И сонеты Шекспира только подливают масла в огонь этих размышлений:

```
шекспировских коллокаций из Бсл: 7438
7438 : 19425  = 0.382
11500 : 19425 = 0.592
шекспировских коллокаций из Мсл: 204
204 : 19425  = 0.010
1368 : 19425 = 0.070
```

Nposit	positSum
5	*54*
10	278
15	530
20	792
25	1321
30	1545
40	2293

50	3075
75	5210
100	8038

Мы, разумеется, не собираемся вот так сразу – даже осторожно – выдвигать предположение, что сонеты написал не Шекспир или не один Шекспир, просто в них меньше шекспировских коллокаций 36-ти пьес, чем во многих других текстах его современников, апокрифах или анонимных произведениях. Возможно, сонеты – особая жанровая форма индивидуального стиля, где меньше повторного.

Кстати, именно компьютерные подсчеты уже выявляли заметную разницу (по другим параметрам) между сонетами и шекспировским драматическим каноном[8], так что сам по себе этот, возможно, не бросающийся в глаза при чтении, но выводимый из строгой математической статистики стилистический зазор между «Шекспиром сонетов» и «Шекспиром 36-ти пьес» не новость.

Скорее новость – обсчет по нашим параметрам знаменитого произведения Роберта Грина, как полагают многие шекспироведы, злейшего завистника Шекспира. Как раз в этом произведении сделан первый прозрачный намек на «потрясателя сцены» (shake scene), собственно этим оно и знаменито (хотя на это произведение стоило бы обратить более пристальное внимание)[9].

«Groats-worth of Witte, bought with a million of Repentance»:

шекспировских коллокаций из Бсл: 4912
4912 : 11359 = 0.**432**
7087 : 11359 = 0.623
шекспировских коллокаций из Мсл: 186
186 : 11359 = 0.*016*
972 : 11359 = 0.**085**

Nposit	positSum
5	*43*
10	*130*
15	400
20	587
25	893
30	1159
40	2128
50	3016
75	4931
100	7742

Эти параметры не намного, но лучше, чем у автора сонетов. Опять-таки опрометчиво было бы заявить, что Грин как автор «На грош ума…» чуть-чуть более Шекспир, чем автор «Шекспировых сонетов». И тут нужны дополнительные исследования, некоторые мы уже осуществили (см. ниже), предварительные результаты выводятся как частный случай применения нашей методики к более коротким отрезкам произведений.

Средние параметры Бсл по 36 пьесам Шекспира по отрезкам текста в 2000 слов такие:

P = **0.875** (0.**871** – по другой методике определения среднего)
P_{use} = 0.996

Для сонетов и «На грош ума…» по Бсл имеем:

№ отрывка (2000)	P сонетов по Бсл	№ отрывка (2000)	P «На грош ума» по Бсл
1	0,478	1	0,486
2	0,549	2	0,564
3	0,567	3	0,540
4	0,552	4	0,541
5	0,565	5	0,575
6	0,519	6	0,545
7	0,521		Ср. **0,542**
8	0,570		
9	0,559		
	Ср. **0,542**		

Удивительное рядом: средняя плотность шекспировских коллокаций совпала до тысячных долей коэффициента!

Очень близка средняя плотность в отрывках «Обесчещенной Лукреции», «Венеры и Адониса» и поэзии Эдварда де Вера («Венера и Адонис» – 0,450 – отличаются от поэзии де Вера вообще всего на 3 тысячных):

№ отрывка (2000)	Р соединенных «Венеры и Адониса» и «Обесчещенной Лукреции» по Бсл	№ отрывка (2000)	Р лирики Э. де Вера по Бсл
1	0.443	1	0.422
2	0.449	2	0.471
3	0.435	3	0.448
4	0.439		Ср. **0.447**
5	0.481	0.443	
6	0.478	0.449	
7	0.450	0.435	
8	0.456	0.439	
9	0.484	0.481	
10	0.489	2253	
11	0.456	Ср. **0.450**	Для «Венеры…»!
12	0.422		
	Ср. **0.457**		

Третья группа произведений составляет отдаленный фон и в какой-то степени может тестировать работу системы. Достаточно показать плотность шекспировских коллокаций по Бсл. Курсивом выделены минимальные значения по каждому произведению, полужирным – максимальные. Для удобства и наглядности в таблице и ниже в текстах из коэффициентов мы будем приводить только цифры после запятой (например, 876 = 0,876), в четырехзначных числах первая цифра будет до запятой (1234 = 1,234).

№ отрывка (2000)	Byron Childe	Byron Juan	Carroll Alice in Wonderland	Chesterton The Ballad of the White Horse	Chesterton The Defendant	Christie Agatha The Mysterious Affair at Styles	Christie Agatha The Secret Adversary	Coleridge Shakespeare Ben Jonson …	Defoe The Further Adventures of Robinson Crusoe	Dickens Bleak House	Dickens David Copperfield
1	401	440	421	444	416	416	386	*320*	**595**	354	426
2	*285*	420	**435**	429	*352*	406	429	334	580	*415*	*414*
3	313	483	387	403	**443**	423	410	384	533	432	432
4	351	475	402	**487**	417	412	431	**400**	565	521	505
5	385	*404*	405	428	401	478	429	397	529	520	505
6	298	**496**	394	391	413	*387*	**460**	399	577	**531**	**518**
7	335	470	369	*389*	373	**511**	*374*	385	591	477	443
8	383	485	349	392	384	480	383	392	518	446	493
9	406	422	382	421	404	480	381	394	537	444	441
10	420	416	393		412	469	414	345	515	509	513
11	383	438	*308*		397	495	399	372	529	461	478
12	**423**	450	356		390	443	417	397	*495*	476	496
Среднее (ср. **431**)	***365***	450	383	420	400	450	409	377	**547**	466	472
Перепад стиля (ср. **105**)	**138**	92	127	98	91	124	86	***80***	100	116	104

Прежде всего стоит заметить, что все произведения относительно однородны по присутствию в разных отрывках шекспировских коллокаций: максимальный перепад значений (138) в «Чайльд Гарольде» Байрона, минимальный – в работе Кольриджа о младших современниках Шекспира (80), средним (нормальным) перепадом можно считать примерно 100 единиц. Таким образом, количество шекспировских коллокаций в художественных текстах – это вовсе не случайное число и является, как можно предварительно судить по этим данным, достаточно стабильным показателем стиля для определенного художественного произведения.

Мы видим, что отдаленный фон художественной литературы дает показатели от 365 (Байрон) до 547 (Дефо). За Дефо вслед идут два произведения Диккенса и одно Агаты Кристи. Средний показатель шекспировской плотности отдаленного фона 431.

С отдаленным фоном более или менее понятно: мы не думаем, что Дефо, Диккенс или Агата Кристи участвовали в создании текстов Шекспира. Наоборот, тексты Шекспира участвовали в создании Дефо и Диккенса. Тут все ясно: прямое и косвенное (через другие, также впитавшие Шекспи-

ра тексты) влияние. Но возьмем произведения Роберта Грина, на которые вроде бы (по сложившимся шекспироведческим представлениям) Шекспир влиять не мог, поскольку в «На грош ума» заявлено о смерти Грина.

Номер отрывка (2000)	*Arbasto*	*Black Books Messenger*	*Ciceronis Amor*	*Debate Betwen Folly*	*Defence Conycatching*	*Dispu-tation*	*Euphues Censure*	*Francescos Fortunes*	*Farewell*	*Mourning Carment*	*Never Too Late*	*Vision*	*Gwydonius*	*Mamillia 1583*
1	492	590	405	**594**	*519*	519	*443*	474	489	472	*405*	484	442	*463*
2	518	611	467	*492*	586	595	479	*460*	486	493	519	528	451	497
3	*489*	**684**	432	496	534	**639**	477	531	443	*457*	514	*411*	453	468
4	520	*587*	**539**		545	582	431	542	469	508	544	528	447	519
5	530		438		588	542	461	561	486	482	436	558	487	521
6	542		*390*		540	*508*	458	516	526	**568**	492	458	504	560
7	566		465		569	558	439	517	**530**	524	460	535	498	539
8	532		467		**623**	568	483	545	*430*	540	436	**604**	457	590
9	525		533			594	**507**	**595**	487	500	512	530	*430*	520
10	**583**		427			603	430	480	527	504	475	554	534	582
11			517				469	483	524	471	491		521	519
12			486				503	545	498	520	**574**		**541**	**601**
Ср. (559)	530	618	464	527	563	571	465	521	491	503	488	519	480	532
Перепад стиля (ср. 131)	94	97	149	102	104	131	64	135	100	110	169	193	111	138

Средний показатель плотности коллокаций пьес Шекспира по 14 произведениям Грина впечатляет: **559**, на 128 единиц выше, чем у отдаленного фона. Если исходить из того, что Шекспир не мог влиять на Грина, то в рамках классических представлений (Шекспир – это один единый автор, а Грин – другой единый автор) остается предполагать колоссальное влияние Грина на Шекспира. (Хотя теоретически можно предположить еще влияние какого-нибудь предшественника и на Грина, и на Шекспира, но на этой, черновой, стадии анализа такими тонкими вариантами влияний можно пренебречь). В этом, на первый взгляд, нет ничего невероятного. Юный будущий автор шекспировского канона зачитывался романами (романсами, точнее переводя, вернее просто калькируя, потому что английский термин romance мало общего имеет с нашими представлением как о романе, так и о романсе) и драмами Роберта Грина. Несколько настораживает уровень стилистического перепада как внутри отдельных произведений Грина (193, 169, 149: максимальные значения), так и между разными произведениями (перепад между средними показателями 154).

(Чтобы проверить гипотезу влияния Грина, нужно составить словники по Грину, аналогичные шекспировским, и сравнить с ними пьесы Шекспира и фона. Но эта проверка уже останется вне рамок данной статьи.)

Однако мы слишком рано нарушили наш собственный запрет на рассмотрение Шекспира как личности. Кто, собственно, дал нам право делать какие-то выводы о возрасте Шекспира и размышлять о том, что он мог читать в юности? Мы даже пока точно не знаем, один это автор или несколько. Мы знаем только, что Шекспир признан автором 36 пьес. Значит, и вернемся к анализу текстов Шекспира. Результаты анализа по Бсл всех пьес Первого Фолио по 2000-м отрывкам тоже представим в виде таблиц.

Номер фрагмента (2000)	*Antony and Cleopatra*	*All is well that ends well*	*As you like it*	*The comedy of errors*	*Coriolanus*	*Cymbeline*	*Midsummer night's dream*	*Hamlet*	*Hen IV1*	*Hen IV2*	*HenV*	*Hen VI 1*	*HenVI 2*
1	815	*789*	994	*817*	*791*	892	817	815	*844*	824	*727*	*739*	845
2	807	889	**1011**	864	811	876	*755*	851	894	974	790	779	910
3	840	806	924	887	847	850	874	*789*	874	863	860	827	**959**
4	823	859	938	**962**	812	829	850	**946**	903	955	770	814	922
5	871	**990**	*793*	915	868	804	868	907	858	*805*	772	836	856
6	828	933	913	933	809	782	774	782	920	879	764	835	*813*
7	*791*	968	944	947	789	847	**885**	860	894	935	787	835	834
8	826	948	972	1077	851	883	774	843	892	890	912	771	875
9	**967**	912	943		900	793		879	920	892	812	**874**	882
10	803	967	986		**919**	802		835	935	861	856	812	899
11	867	932	905		842	*722*		836	957	887	**934**	964	891
12	869	1016			856	**995**		883	**978**	905	827		899
13					865	827		805		**992**	830		997
14					803	1170		871					
15								870					
Ср.	842	908	**938**	903	840	839	825	851	906	897	819	**812**	882
Перепад стиля	176	201	218	148	**128**	**273**	130	157	134	187	207	135	146

Номер фрагмента (2000)	*Hen VI 3*	*Hen VIII*	*Julius Caesar*	*John*	*Lear*	*Love's labor's lost*	*Measure for measure*	*The merry wives of Windsor*	*Macbeth*	*The merchant of Venice*	*Much ado about nothing*	*Othello*	*Romeo and Juliet*
1	935	804	913	859	812	816	901	890	*747*	882	999	878	859
2	872	*751*	*872*	800	872	833	880	850	797	888	966	859	804
3	837	865	911	*753*	875	**891**	**1029**	971	798	906	**1001**	827	818
4	899	918	920	898	896	806	856	969	796	928	*927*	*813*	840
5	*828*	916	904	812	802	744	*829*	808	830	911	948	858	878
6	**984**	848	939	893	867	775	911	**973**	**842**	*870*	957	963	898
7	844	862	885	900	788	*702*	905	966	835	**1018**	970	961	*734*
8	936	832	921	**906**	790	779	891	815	796	921	945	927	923
9	869	866	927	819	847	878	926	922	959	933	973	888	**943**
10	875	**962**	**978**	872	*772*	814	935	824		962	955	**968**	883
11	864	891			**887**	993	936	966		1100	1084	927	855
12	911	867			874							935	870
13					911							889	
Ср.	888	865	917	851	846	**804**	909	899	805	922	**964**	899	859
Перепад стиля	176	171	106	153	115	189	200	167	95	148	**74**	155	**209**

Номер фрагмента (2000)	*Richard II*	*Richard III*	*Twelfth night*	*The taming of the shrew*	*The tempest*	*Timon of Athens*	*Titus Andronicus*	*Troilus and Cressida*	*The two gentlemen of Verona*	*The winter's tale*		
1	792	868	*865*	855	780	816	*756*	840	893	*793*		
2	827	921	929	882	743	878	848	*700*	940	826		
3	*773*	912	950	910	816	810	797	708	924	895		
4	845	955	919	**976**	801	**894**	873	776	*877*	867		
5	**980**	945	932	841	853	823	839	812	905	827		
6	819	967	963	829	801	*693*	859	866	893	849		
7	868	**1027**	**988**	*828*	*707*	728	852	850	992	837		
8	875	956	959	973	**879**	797	863	795	978	833		
9	859	880	962	960		751	**946**	**928**	1000	894		
10	936	983	972	930			813	797		**944**		
11	884	*820*		924				850		834		
12		873						751		904		
13		899						823		850		
14		858										
15		810										
Ср.	860	919	944	901	798	799	845	807	925	858	Ср. для 36 пьес	**871**
Перепад стиля	207	207	123	148	172	201	190	**228**	115	151	Ср. для 36 пьес	**165**

(Подчеркиванием выделены последние отрывки произведений меньше 1000 слов, где плотность коллокаций может резко увеличиваться чисто математически.)

Пока мы можем определенно сказать, что средний перепад стиля между отрывками в 2000 слов внутри шекспировских произведений для 36 пьес равен 158 против 105 в отдаленном фоне. Даже у очень разношерстного Грина средний уровень стилистического перепада ниже (131). Причины этого могут быть как чисто математические (у Шекспира, естественно, больше шекспировских коллокаций, а значит, и больше колебания числа этих коллокаций), так и историко-лингвистические (период становления английского языка в эпоху Шекспира, с одной стороны, и более устоявшийся язык в более поздние времена, с другой стороны). В любом случае, исходя лишь из анализа по Бсл как пьес в целом, так и их отрывков по 2000 слов, мы пока не находим оснований сокращать шекспировский корпус.

Однако, проанализировав полученные данные, мы можем подозревать участие соавторов в некоторых пьесах. Например, начальные 2000 слов «Тита Андроника» резко выбиваются по коэффициенту шекспировских коллокаций, но не так резко, чтобы наши подозрения превратились сразу в нечто большее, хотя ученые сейчас практически не сомневаются в участии в пьесе нешекспировской руки, точнее руки Джорджа Пила. То есть, именно говоря об этом произведении, мы можем в какой-то степени проверить адекватность методики. Если отрывки, где у нас подозревается соавторство, совпадут с отрывками неШекспира, полученными по другим методикам, значит, методики первый тест на адекватность выдержали. Если не совпадут, потребуется коррекция, правда, чьей методики (нашей или иных), заранее сказать трудно.

Большинство исследователей, занимавшихся проблемой атрибуции «Тита Андроника» и признававших факт соавторства, сходятся во мнении, что Дж. Пил написал первый акт, первую сцену второго акта и первую сцену четвертого акта. Тесты, проведенные с помощью компьютеров, это подтверждают[10]. Все эти исследования подсчитывали самые разные, но частные показатели стиля, начиная от соотношения мужских и женских окончаний стихов и кончая количеством более чем двусложных слов. Проведенный с помощью практического полного словника шекспировских коллокаций (133 тысячи единиц) анализ пьесы, последовательно разбитой на 2000-е и 1000-е отрывки, приводит к частично похожим выводам:

Номер фрагмента (2000)	Акт, сцена «Тита Андроника»	Плотность 2000-х	Плотность 1000-х	Номер фрагмента (1000)
1	**1 акт**	*756*	807	1
			807	2
2	**1 акт**	848	868	3
			905	4
3	**1 акт** (248 слов) **2 акт 1 сц.** 752	797	869	5
	2.1. (295 слов) 2.2.		*773*	6
4	2.2. 2.2.	873	912	7
			905	8
5	2.2. 2.2. + 3.1. (207 слов)	839	841	9
			888	10
6	3 акт 2 сц. и	859	930	11
	4 акт 1 сц. (1012 слов)		844	12
7	**4 акт 1 сц.** (22 слова)	852	901	13
			884	14
8		863	902	15
			882	16
9		**946**	975	17
			1007	18
10		813	830	19
			855	20
	Среднее:	**845** (по 36 пьесам **871**)	**897** (по 36 пьесам **901**)	

Хотя мы видим, что наши 2000-е отрывки, естественно, не совпадают с делением по сценам, какие-то предварительные выводы сделать можно. Например, можно однозначно подтвердить, что текст первой половины 1 акта (чуть меньше) имеет показатель плотности коллокаций наименее шекспировский. Если признать, что соавтором был Пил, то самое начало в 2000 слов написал именно он. Затем 2000 слов идут со вполне шекспировскими показателями плотности. Возможно, вторую половину 1 акта все-таки написал главный Шекспир (условно пишем так, потому что при некоторых выводах из стилистических исследований и самого Пила логично считать Шекспиром). Самые последние 248 слов 1 акта и 1 сцена 2 акта (1047 слов) входят в нашу третью порцию по 2000 слов, которая имеет показатель плотности ниже среднего по пьесе и ниже среднего по 36 пьесам.

Вполне можно предположить авторство Пила. А вот последний отрывок, обычно приписывающийся Пилу, мы по анализу 2000-х отрывков атрибутировать не можем, потому что он занимает лишь половину нашего шестого отрезка текста и самое начало седьмого. Однако и по анализу 1000-х отрывков картина получается неоднозначная. С одной стороны, наш 12-й отрывок, в который почти полностью вмещается подозреваемая на авторство Пила 1 сцена 4 акта, достаточно резко отличается от предыдущего, 11-го и последующего, 13-го. С другой стороны, в целом 12-й отрывок в 1000 слов вполне вписывается в шекспировские показатели. Если этот 12-й отрывок считать нешекспировским, то и 9, и 19 отрывок тоже нужно считать нешекспировскими по этим показателям.

Однако в целом наши результаты не противоречат результатам, полученным по другим методикам, да и, конечно, сами методики, которые сейчас применяются, наш подход никоим образом не отрицает. Главное отличие предлагаемой методики состоит в глобальности стилистического анализа: это **сплошной** анализ текста, который раньше можно было проводить только в ручном режиме, а значит, охватывались только незначительные отрезки текста в единицу времени, сопоставимую с творческой жизнью исследователя.

ПРИМЕЧАНИЯ

[1] *Bolton W.* The Bard in Bits: Electronic Editions of Shakespeare and Programs to Analyze Them // Computers and the Humanities. 1990. Vol. 24. №. 4 (Aug.). P. 275–287.

[2] *Lowe D., Matthews R.* Shakespeare Vs. Fletcher: A Stylometric Analysis by Radial Basis Functions // Computers and the Humanities. 1995. Vol. 29. № 6 (Dec.). P. 449–461.

[3] *Elliott W.E., Valenza R.J.* A Touchstone for the Bard // Computers and the Humanities. 1991. Vol. 25. № 4 (Aug.). P. 199–209.

[4] *Eftekhari A.* Fractal geometry of texts: An initial application to the works of Shakespeare // Journal of Quantitative Linguistics. 2006. Vol. 13. № 2–3. P. 177–193.

[5] *Simonton D.K.* Lexical Choices and Aesthetic Success: A Computer Content Analysis of 154 Shakespeare Sonnets // Computers and the Humanities. 1990. Vol. 24. № 4 (Aug.). P. 251–264.

[6] http://research.cs.wisc.edu/niagara/data/shakes/shaksper.htm

[7] *Ward E.Y.* Elliott and Robert J. Valenza. Shakespeare's Vocabulary: Did it Dwarf All Others? // Stylistics and Shakespeare's Language / Eds Mireille Ravassat and Jonathan Culpeper. London; New York, 2011. P. 34–41.

[8] *Zhang Katherine T., Zhang Zhiyi.* Shakespearean Sonnets versus Shakespearean Canon // Journal of Quantitative Linguistics. 2010. Vol. 17. № 2. P. 81–93.

[9] *Пешков И.В.* Почему Роберт Грин за грош ума (остроумия) каялся на миллион, или Львиная природа авторства // Бестиарий в словесности и изобразительном искусстве. М., 2012. С. 97–126.

Peshkov I.V. Pochemu Robert Grin za grosh uma (ostroumija) kajalsja na million, ili L'vinaja priroda avtorstva // Bestiarij v slovesnosti i izobrazitel'nom iskusstve. Moscow, 2012. P. 97–126.

[10] *MacDonald P. Jackson.* Studies in Attribution: Middleton and Shakespeare. Salzburg, 1979. P. 147–153; *Tarlinskaja M.* Shakespeare's Verse: Iambic Pentameter and the Poet's Idiosyncrasies. New York, 1987. P. 121–124; *Vickers B.* Shakespeare, Co-Author: A Historical Study of Five Collaborative Plays. Oxford, 2002. P. 219–239.