

## **ПРОБЛЕМА «РОМЕН ГАРИ – ЭМИЛЬ АЖАР»: АТРИБУЦИЯ РОМАНОВ, ОПУБЛИКОВАННЫХ ПОД ПСЕВДОНИМОМ ЭМИЛЬ АЖАР**

*Работа представлена кафедрой романской филологии  
Санкт-Петербургского государственного университета.*

*Научный руководитель – доктор филологических наук, профессор М. А. Марусенко*

**В статье рассматриваются результаты атрибуции трех романов, опубликованных под псевдонимом Эмиль Ажар, с помощью метода распознавания образов, который предусматривает применение многомерной классификации и описания индивидуального авторского стиля в синтаксическом аспекте. Впервые данный метод применяется на материале современного французского языка.**

**The article considers the results of attribution of the three novels published under the pseudonym Emile Ajar by means of the method of pattern recognition. This method provides application of multivariate classification and description of the author's individual style in the syntactic aspect. For the first time the method is applied by the material of the modern French language.**

Ромен Гари (1914–1980), являясь одним из крупных французских писателей XX в., привлекает внимание как широкого круга читателей, как и исследователей прежде всего своей экспериментальной деятельностью в области художественной литературы.

Наибольший интерес вызывает проблема авторства романов, написанных под

псевдонимом Эмиль Ажар. В период с 1974 по 1980 г. литературными критиками и сотрудниками издательских домов «Галлимар» и «Меркюр де Франс» не ставилась под сомнение непринадлежность четырех произведений Э. Ажара Ромену Гари, хотя они вычитывали и публиковали параллельно произведения как Р. Гари, так и Э. Ажара.

Романы, написанные Ажаром, считали творением Р. Кено, Л. Арагона и М. Турнье, Ж. Ланцмана, работой коллектива авторов. Затем – П. Павловича, племянника Р. Гари.

После выхода в свет романа П. Павловича «Человек, которому верили»<sup>1</sup>, а также посмертной публикации Р. Гари «Жизнь и смерть Эмиля Ажара»<sup>2</sup> литературный мир потряс тот факт, что Р. Гари и Э. Ажар являются одним и тем же человеком, к тому же Р. Гари оказался единственным писателем, получившим два раза Гонкуровскую премию.

Работы по атрибуции псевдонимных романов Э. Ажара сводятся к одному диссертационному труду 1998 г., проведенному польским исследователем А. Павловски, который использовал модель временных рядов<sup>3</sup>. Результатом работы стала атрибуция четырех произведений Э. Ажара Ромену Гари. При этом исследователь не учитывал жанровое несоответствие романов Э. Ажара и привлек в качестве потенциальных авторов ряд современников Р. Гари, основываясь только на газетных заметках того времени, появившихся до установления того факта, что в данной литературной «мистификации» участвовал П. Павлович, исполнявший роль Э. Ажара.

Таким образом, по результатам истории вопроса и исследования архивных документов мы сталкиваемся с сосуществующими противоречащими точками зрения, разрешение которых должно быть достигнуто путем проверки различных гипотез. При этом нулевую гипотезу можно формулировать следующим образом: ( $H_0$ ) – все тексты романов Э. Ажара полностью принадлежат Р. Гари. Соответственно, альтернативные гипотезы будут выглядеть так: ( $H_a^1$ ) – тексты романов Э. Ажара являются произведением П. Павловича и ( $H_a^2$ ) – тексты романов Э. Ажара являются совместным произведением Р. Гари и П. Павловича с определенной долей участия каждого из них. Естественно, что вопрос об атрибуции того или иного романа Э. Ажара решается в каждом случае индивидуально.

Для практической реализации этой процедуры необходимо располагать минимум двумя объектами, принадлежащими к двум разным классам, известным заранее<sup>4</sup>. В нашем исследовании такими классами являются классы  $\Omega_1$  (Gary) и  $\Omega_2$  (Pavlowitch).

Исходя из требования соблюдения синхронии и жанрово-стилистической однородности текстов, корпус текстов, используемых для определения координат эталонов классов, имеет следующую структуру (табл. 1):

Таблица 1

Состав и объекты априорных классов

Класс	Код	Произведения	Год издания	Объем авторской речи (N≈)
$\Omega_1$ (Gary)	1G	Gary R. Au-delà de cette limite votre ticket n'est plus valable.	1975	5233
	2G	Gary R. Clair de femme.	1977	
	3G	Gary R. Les Cerfs-volants.	1980	
$\Omega_2$ (Pavlowitch)	1P	Pavlowitch P. Un autre monde.	2004	5200
	2P	Pavlowitch P. Tom.	2005	

Относительно сформированных априорных классов будет производиться классификация многомерных объектов (тек-

стов), написанных под псевдонимом Эмиль Ажар, подлежащих распознаванию (табл. 2):

Таблица 2

Атрибутируемые объекты

Код	Наименование	Год издания	Объем авторской речи (N≈)
1A	Gros-Câ lin	1974	1449
2A	La Vie devant soi	1975	2394
3A	L'Angoisse du roi Salomon	1979	3078

Из текстов каждого автора, входящих в табл. 1, были сделаны случайные предварительные выборки объемом 200 предложений, для каждой из которых были определены значения 54 параметров из исходного набора<sup>5</sup>. В соответствии с используемой процедурой атрибуции следующим этапом является выделение из априорного набора параметров, способных разделять априорные классы. Результаты проверки

существенности разности средних для двух априорных классов с помощью *t*-критерия Стьюдента (формула 1) приведены в табл. 3.

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{1/2}} \quad (1)$$

**Таблица 3**

**Проверка статистической гипотезы существенности разности двух средних**

Параметр	Ω(Gary)/ Ω(Pavlowitch)	Параметр	Ω(Gary)/ Ω(Pavlowitch)
	<i>t</i> <sub>набл.</sub>		<i>t</i> <sub>набл.</sub>
X01	1,789	X28	<b>3,262</b>
X02	<b>2,093</b>	X29	<b>2,336</b>
X03	<b>2,944</b>	X30	<b>3,753</b>
X04	0,096	X31	<b>6,160</b>
X05	1,642	X32	0,080
X06	<b>3,506</b>	X33	1,930
X07	<b>3,052</b>	X34	1,372
X08	<b>3,038</b>	X35	1,016
X09	1,418	X36	0,936
X10	1,000	X37	1,076
X11	<b>5,193</b>	X38	0,161
X12	1,926	X39	0,453
X13	1,733	X40	0,980
X14	1,807	X41	0,112
X15	1,485	X42	0,814
X16	1,662	X43	0,902
X17	1,160	X44	1,306
X18	0,535	X45	1,893
X19	<b>6,971</b>	X46	1,517
X20	<b>4,382</b>	X47	1,260
X21	<b>4,816</b>	X48	0,000
X22	1,555	X49	0,539
X23	<b>4,292</b>	X50	0,372
X24	<b>2,842</b>	X51	1,669
X25	<b>3,386</b>	X52	1,556
X26	1,740	X53	1,477
X27	0,537	X54	1,867

Из табл. 3 видно, что статистически значимой является разность средних для параметров X02, X03, X06, X07, X08, X11, X19,

X20, X21, X23, X24, X25, X28, X29, X30, X31, для которых наблюдаемые значения *t*-критерия превышают критические (*t*<sub>0,05</sub> = 1,96).

На втором этапе из подмножества шестнадцати параметров удалось выделить подмножество из трех параметров, удовлетворяющих требованиям сильной корреляции с другими параметрами совокупности и способности к разделению объектов на классы. Таким образом, рабочий словарь параметров распознающей системы по атрибуции романов Э. Ажара включает 3 диагностических параметра: X23 – число предлогов; X24 – число союзов; X28 – число прямых дополнений.

Поскольку произведения, входящие в корпус текстов (табл. 1), значительны по объему и сплошное статистическое обследование не является целесообразным, следующим этапом процедуры определения координат эталонов классов является определение объема выборки<sup>6</sup> (формула 2), результаты которого представлены в табл. 4–5.

$$n = \frac{N}{1 + \left(\frac{V_{\bar{x}}}{V}\right)^2} N, \quad (2)$$

где  $V = \frac{s}{\bar{x}}$ ,  $V_{\bar{x}} = \frac{V}{\sqrt{n}} \sqrt{1-f}$ ,

$s$  – выборочное среднее квадратичное отклонение,  $f$  – доля отбора,  $\bar{x}$  – выборочное среднее значение признака для совокупности;  $V_{\bar{x}}$  – относительная стандартная ошибка, которая показывает, сколько процентов от истинного значения составляет стандартная ошибка оценки.

При  $V_{\bar{x}} = 0,05$  стандартная ошибка составляет только 5% величины оцениваемого параметра.

Анализ данных таблиц показывает, что в каждом классе необходимый объем выборки для различных параметров варьирует в широких пределах. Например, для объекта 1P для параметра X23 необходим минимальный объем выборки ( $n = 1023$ ), а для определения значения параметра X24 с  $V_{\bar{x}} = 0,05$  необходима выборка объемом в 1861 предложение. Произведя случайную выборку в 1861 предложение, мы с доста-

Таблица 4  
Определение объема выборки для объектов класса  $\Omega_1$  (Gary)

Код произведения	Параметр	$\bar{x}_i$	$\sigma_i$	$n_{(V_{\bar{x}}=0,05)}$
1G	X23	2,970	2,588	688
	X24	1,090	1,443	<b>1034</b>
	X28	1,640	1,432	690
N=1680				
2G	X23	1,930	1,833	589
	X24	0,810	0,961	<b>695</b>
	X28	1,430	1,328	578
N=1026				
3G	X23	3,600	2,741	658
	X24	1,040	1,127	<b>1053</b>
	X28	1,720	1,386	715
N=2527				

Таблица 5  
Определение объема выборки для объектов класса  $\Omega_2$  (Pavlovitch)

Код произведения	Параметр	$\bar{x}_i$	$\sigma_i$	$n_{(V_{\bar{x}}=0,05)}$
1P	X23	1,600	1,596	1023
	X24	0,450	0,783	<b>1861</b>
	X28	1,870	0,991	1214
N=3100				
2P	X23	2,340	2,903	1112
	X24	0,530	0,822	<b>1339</b>
	X28	0,870	0,960	989
N=2100				

точной точностью определим значение параметра X24, а значения остальных параметров будут измерены с заведомо избыточной точностью. Результаты обработки данных, полученных методом случайной выборки, представляющие собой координаты эталонов классов на нулевой итерации, приведены в табл. 6.

**Таблица 6**  
Координаты эталонов классов на 0-й итерации

Параметр	Класс			
	$\Omega^0(\text{Gary})$		$\Omega^0(\text{Pavlowitch})$	
	$\bar{X}_i$	$\sigma_i$	$\bar{X}_i$	$\sigma_i$
X23	2,281	2,350	1,525	1,712
X24	0,856	1,139	0,401	0,714
X28	1,431	1,284	0,779	0,898
	N=2782		N=3200	

Следующим этапом процедуры распознавания является определение координат распознаваемых объектов (трех романов Э. Ажара). Статистические характеристики распределений параметров для всех атрибутируемых объектов приведены в табл. 7.

**Таблица 7**  
Координаты атрибутируемых объектов

Объект	Параметр	$\bar{X}_i$	$\sigma_i$	N
1A	X23	3,090	3,198	2394
	X24	1,250	1,520	
	X28	1,430	1,451	
2A	X23	2,160	1,873	1449
	X24	1,710	1,578	
	X28	1,780	1,433	
3A	X23	2,150	2,492	3078
	X24	1,160	1,376	
	X28	1,570	1,533	

В соответствии с применяемым алгоритмом атрибуция производится в два этапа: при помощи детерминированного алгоритма распознавания, затем, после остановки последнего, при помощи вероятностного алгоритма распознавания. В описываемом эксперименте остановка детерминированного алгоритма распознавания произошла на 0-й итерации.

После остановки детерминированного алгоритма все три атрибутируемых объекта оказались не отнесенными ни к одному

априорному классу. Реализация вероятностного алгоритма распознавания предусматривает преобразование исходной матрицы данных в матрицу евклидовых расстояний между объектами, подлежащими распознаванию, и эталонами априорных классов (табл. 8).

**Таблица 8**  
Матрица евклидовых расстояний между классами и неатрибутированными объектами

	$d(X_j, \Omega_i)$		
	Объект		
Класс	1A	3A	4A
$\Omega_1$	0,900	0,931	0,359
$\Omega_2$	1,986	1,766	1,262

Следующим этапом процедуры вероятностного распознавания является формирование решающего правила, при котором происходит приписывание вероятности или достоверности каждому решению, что делает возможным решать любую задачу отнесения объектов к априорным классам. Вероятность принадлежности j-го объекта к i-му классу (табл. 9) вычисляется на основе матрицы евклидовых расстояний между объектами и априорными классами по следующей формуле:

$$P_{ji} = \frac{1}{d_{ji}} \left( \sum_k \frac{1}{d_{jk}} \right)^{-1}, \quad (3)$$

где  $d_{ji}$  – расстояние между j-м объектом и i-м классом,  $d_{jk}$  – расстояние между j-м объектом и остальными классами классификации<sup>7</sup>.

**Таблица 9**  
Матрица вероятностей принадлежности объектов априорным классам  $P(X_i \in \Omega_j)$

	$P(X_i \in \Omega_j)$		
	Объект		
Класс	1A	2A	3A
$\Omega_1$	0,678	0,655	0,778
$\Omega_2$	0,322	0,345	0,222

При наличии двух априорных классов решающее правило может быть сформулировано как:

$$X_i \in \Omega_2, P(X_i \in \Omega_2) > \sum_{n=1}^2 P(X_i \in \Omega_n).$$

Таким образом,

$$X_i \in \Omega_j, P(X_i \in \Omega_j) > 0,5,$$

т. е. при данном решающем правиле пороговое значение вероятности принадлежности объекта к одному из априорных классов должно превышать 0,5. По данным табл. 9, классу  $\Omega$  (Gary) могут быть атрибутированы все три объекта: A1 (Gros-Cvlin), A2 (La Vie devant soi) и A3 (L'Angoisse du roi Salomon).

Правильность и правомочность полученных результатов после проведения вероятностной атрибуции необходимо проверить, проведя оценку однородности априорного класса по сравнению с априорным классом. Критерием однородности служит средний квадрат расстояния между объектами  $\bar{d}(\Omega_N)$ . Построим матрицу евклидовых расстояний между объектами априорного класса (табл. 10).

В качестве мер близости между объектами одного класса  $\Omega_p, p = 1, 2, \dots, m$  обычно используют величину, называемую среднеквадратичным разбросом объектов внутри класса  $\Omega_p$ :

$$\bar{S}(\Omega_p) = \left[ \frac{2}{k_p} \frac{1}{k_p - 1} \sum_{k=1}^{k_p} \sum_{l=1}^{k_p} d^2(\omega_{pk}, \omega_{pl}) \right]^{1/2}. \quad (4)$$

Качество классификации считается тем выше, чем теснее расположены объекты внутри класса.

Таблица 10

Матрица евклидовых расстояний между объектами априорного класса  $\Omega_1$  (Gary)

Объект	$d(X_i, X_j)$		
	Объект		
	1G	2G	3G
1G	0,000	1,379	0,243
2G	1,379	0,000	1,541
3G	0,243	1,541	0,000

С помощью формулы 4 вычислим среднеквадратичный разброс для объектов класса  $\Omega_1$  (Gary):  $\bar{S}(\Omega_3) = 1,523$ .

Построим следующую матрицу данных, объединив исходные данные трех объектов, составляющих класс  $\Omega_1$  (Gary), и всех трех атрибутируемых Гари объектов по вероятностной атрибуции (табл. 11).

Таблица 11

Матрица евклидовых расстояний между объектами апостериорного класса  $\Omega_{Gary, Ajar}$

Объект	$d(X_i, X_j)$					
	Объект					
	1G	2G	3G	1A	2A	3A
1G	0,000	1,379	0,243	0,801	1,385	1,008
2G	1,379	0,000	1,541	1,150	1,389	0,914
3G	0,243	1,541	0,000	1,499	1,552	1,218
1A	0,801	1,150	1,499	0,000	0,789	0,318
2A	1,385	1,389	1,552	0,789	0,000	0,950
3A	1,008	0,914	1,218	0,318	0,950	0,000

С помощью формулы (4) вычислим среднеквадратичный разброс для объектов апостериорного класса  $\Omega_{Gary, Ajar}$ :  $\bar{S}(\Omega_{15}) = 1,148$ .

Поскольку  $\bar{S}(\Omega_3) > \bar{S}(\Omega_{Gary, Ajar})$ , можно сделать вывод о том, что полученная классификация объектов улучшилась и близка к естественной, что подтвердило результаты работы вероятностного алгоритма и принадлежность априорному классу трех атрибутируемых объектов.

Представим полученные данные в виде матрицы решений, где 0 – отказ от распознавания, 1 – положительное решение по детерминированному алгоритму атрибуции, 1\* – положительное решение по вероятностному алгоритму атрибуции (табл. 12).

Таблица 12

Матрица решений

Код	Наименование	$\Omega_1$ (Gary)	$\Omega_2$ (Pavlowitch)
1A	Gros-Calin	1*	0
2A	La Vie devant soi	1*	0
3A	L'Angoisse du roi Salomon	1*	0

В результате проведенного исследования была подтверждена гипотеза ( $H_0$ ): тексты романов Э. Ажара полностью принадлежат Р. Гари. Таким образом, атрибутируемые объекты были приписаны априорному классу  $\Omega_1$  (Gary) и был сформирован апостериорный класс  $\Omega_{Gary, Ajar}$ , показавший высокую степень однородности.

Были опровергнуты сформулированные во второй главе гипотезы ( $H_a^1$ ) и ( $H_a^2$ ), согласно которым тексты романов Э. Ажара являются произведением П. Павловича ( $H_a^1$ ) или совместным произведением Р. Гари и П. Павловича с определенной долей участия каждого из них ( $H_a^2$ ).

Представляется возможным отойти от рассмотрения Р. Гари и Э. Ажара как одного человека, творившего под двумя разными псевдонимами, и разграничить такие термины, как «писатель» и «автор». Данное разграничение предложено французским литературоведом Ф. Вернье. Так, писателем она предлагает признать индивидуума, писательство для которого является профессиональной деятельностью. У такого человека есть личная, творческая, социальная история, к которой мы можем обратиться в литературоведческом исследо-

вании. Автора же можно рассматривать как личность, которая раскрывается путем исследования конкретного литературного наследия, стилистики, интенциональности его произведений<sup>8</sup>. Если принять данную точку зрения, то мы можем рассматривать Р. Гари и Э. Ажара как двух разных авторов, объединенных одним писателем – Роменом Гари.

Тот факт, что все результаты получены путем вероятностной атрибуции, может быть интерпретирован следующим образом: Р. Гари, скрывавший свое авторство, намеренно старался изменить стиль с тем, чтобы его мистификация не была сразу же разгадана. Именно этим может объясняться отсутствие результатов детерминированной атрибуции. Но поскольку писатель не может значительно изменять характеристики латентных синтаксических структур при создании больших объемов текста, то вероятностная атрибуция смогла однозначно указать на истинного автора.

Результаты исследования служат еще одним показателем того, что характеристики индивидуального авторского стиля носят объективный характер и не зависят от воли пишущего.

#### ПРИМЕЧАНИЯ

<sup>1</sup> Pavlowitch P. L'Homme que l'on croyait. P.: Fayard, 1981.

<sup>2</sup> Gary R. Vie et mort d'Emile Ajar. P.: Gallimard, 1981.

<sup>3</sup> Pawiowski A. Sériés temporelles en linguistique avec application a l'attribution de textes: Romain Gary et Emile Ajar. P.: H. Champion, 1998.

<sup>4</sup> Марусенко М. А. Атрибуция анонимных и псевдонимных литературных произведений методами распознавания образов. Л.: Изд-во ЛГУ, 1990. С. 129.

<sup>5</sup> Там же. С. 71–74.

<sup>6</sup> В поисках потерянного автора: Этюды атрибуции / М. А. Марусенко, Б. Л. Бессонов, Л. М. Богданова и др.; Под ред. М. А. Марусенко. СПб.: Фил. фак-т СПбГУ, 2001. С. 14–15.

<sup>7</sup> Терехина А. Ю. Анализ данных методами многомерного шкалирования. М.: Наука, 1986. С. 153.

<sup>8</sup> Vernier F. L'ange de la thйorie // Paragraphes. Montrйal: Presses universitaires de Montrйal, 2004. P. 120–122.