

С. Г. ЧЕБАНОВ

О ПОДЧИНЕНИИ РЕЧЕВЫХ УКЛАДОВ «ИНДО-ЕВРОПЕЙСКОЙ»
ГРУППЫ ЗАКОНУ ПУАССОНА

(Представлено академиком А. Н. Колмогоровым 14 II 1946)

1. В процессе стадийного развития языков на каждом определенном историческом этапе коллективным творчеством человеческих обществ вырабатываются речевые уклады.

Одной из важнейших статистических характеристик того или иного уклада речи является распределение слов по числу входящих в них слогов*. Характеристика эта обладает большой устойчивостью в пределах одного текста или нескольких близких по структуре текстов и, наконец, даже в пределах некоторой языковой группы.

Например, при статистическом обследовании текстов из Истории Тита Ливия, общим объемом в 20 128 слов, обнаружено такое распределение слов по числу слогов:

Таблица 1

Число слогов в слове	1	2	3	4	5	6	7	Всего
Число слов	4493	6089	5608	2995	770	135	38	20128
То же в %	22,3	30,3	27,9	14,8	3,8	0,7	0,2	100

В отдельных, меньшего объема, выборках из тех же глав Истории Тита Ливия обнаруживаем следующее распределение слов по числу слогов (в процентах):

Таблица 2

Г л а в ы	Число слогов в слове (n)							Всего слов
	1	2	3	4	5	6	7	
XXX, 32—33	23,1	29,1	27,9	15,5	3,4	0,8	0,2	587
XXIX, 25—29, 35	21,4	30,0	29,3	14,4	3,9	0,8	0,2	2055
XXV, 30—35								
XXVI, 4—8								
XXVI, 9—12	22,3	30,5	28,6	14,2	3,6	0,6	0,2	5133

Мы видим, что уклад речи, в смысле распределения слов по числу слогов, в каждой из этих выборок мало отличается от среднего для всех изученных глав Тита Ливия.

* В настоящем исследовании элементом речи принимается слово, как морфологически вполне определенная частица речи, имеющая более или менее выраженное смысловое значение. Встречающиеся неясности и неточности в определении того или иного слова игнорируются при оперировании с множествами, если число этих неясностей недостаточно велико, чтобы повлиять на результаты подсчетов. Это игнорирование требует строгого и точного учета всех сомнительных случаев.

Аналогичные примеры могут быть приведены и из других языков «индо-европейской» группы (различных эпох), но размеры настоящей статьи позволяют привести лишь следующие, наиболее характерные. Так, имеем в «ново-немецком» языке:

Лессинг, «Лаокоон» (1)

Таблица 3

О т д е л ы	Число слогов в слове (n)						Всего слов
	1	2	3	4	5	6	
Предисл. и I	46,3	33,3	14,7	4,4	1,0	0,3	2307
II и III	47,8	32,4	13,1	5,4	1,1	0,2	2648
IV	48,6	31,0	13,7	5,0	1,4	0,3	3247
V, VI	47,4	34,1	12,9	4,6	0,8	0,2	3375
VII, VIII, IX, X	47,7	34,4	12,4	4,2	1,1	0,2	4107
XI, XII, XIII	51,5	30,0	13,5	3,9	0,9	0,2	3399
XIV, XV, XVI до стр. 104 (вкл.)	47,5	32,8	13,5	4,7	1,3	0,2	1225
Общее	48,5	32,3	13,1	4,7	1,2	0,2	20308

Интересно отметить, что в германских языках общая форма уклада мало изменилась на протяжении 1200 лет.

Таблица 4

Век.	Т е к с т	Число слогов в слове						Всего слов
		1	2	3	4	5	6	
VI	Улфила (2)	43,6	37,3	13,4	4,9	0,8	0,0	2046
IX	Хелианд (3)	47,0	36,7	13,5	2,5	0,3	0,0	3350
X	Татиан (4)	44,1	37,2	13,9	3,9	0,7	0,2	1523
XIII	Эйке ф. Ренгов (5)	48,5	36,5	12,0	2,8	0,2	0,0	1513
XVIII	Лессинг (1)	48,5	32,3	13,1	4,7	1,2	0,2	20308
XIX	Гейгер (6)	47,1	33,4	12,6	5,5	1,1	0,3	2120

Сохранение устойчивых форм укладов отмечается также и в «народных» типах речи, например, французских «patois».

Так, для «аквитанской» группы «patois» XVI—XVII вв. имеем:

Таблица 5

P a t o i s	Число слогов в слове (n)					Всего слов
	1	2	3	4	5	
Charente (7)	53,6	32,2	11,4	1,6	0,2	183
Sarladais (8)	54,2	32,8	10,5	1,4	0,1	540
N. Garonne (9)	54,8	32,3	11,1	1,6	0,2	521
Limousin (10)	52,2	33,3	11,4	1,9	0,2	464
St. Vrieux (11)	54,8	31,8	11,3	1,9	0,2	505

Всего в течение 20 лет нами было обследовано 127 различных речевых форм «индо-европейской» группы с аналогичными результатами.

2. Распределение слов по числу слогов численно выражается частотами $P(n)$, с которыми в данном тексте встречаются слова из n слогов. Например, в рассмотренных выше главах Тита Ливия $P(1) = 0,223$, $P(2) = 0,303$, $P(3) = 0,279$ и т. д.

Возникает естественный вопрос, нельзя ли указание всех частот $P(n)$ для $n = 1, 2, 3, 4, \dots$ заменить указанием одной сводной характеристики, а именно среднего числа слогов в слове

$$\alpha = \sum_n n(Pn). \quad (1)$$

Для выяснения этого вопроса на прилагаемом графике нанесены

Таблица 6

Таблица согласия средней (ν) и центральных моментов (μ_2 и μ_3)

№ п/п	Текст	Язык	Nz						N	m_2	ν	μ_2	μ_3
			1	2	3	4	5	6					
1	Parzival ⁽¹²⁾	Верхн.-немецк.	1823	849	194	37	—	—	2903	2,60	0,47	0,44	0,46
2	Nieblungen ⁽¹³⁾	»	1945	1083	245	50	—	—	3323	2,84	0,52	0,53	0,34
3	d. Goxrud ⁽¹⁴⁾	Иранский	719	573	207	55	12	—	1566	3,87	0,76	0,75	0,68
4	Gobhila-putra ⁽¹⁵⁾	Санскрит	155	230	183	101	42	19	730	8,44	1,60	1,68	1,60
5	Brocans hymnen ⁽¹⁶⁾	Ст. ирландск.	525	362	79	9	—	—	975	6,50	0,90	0,47	0,51
6	Vie de St. Léger ⁽¹⁷⁾	Ст. французск.	537	218	75	4	—	—	834	5,47	2,58	0,46	0,40
7	d. Pásto ⁽¹⁸⁾	Афганск.	495	218	49	16	4	—	782	6,50	2,77	0,49	0,74
8	Ullila ⁽²⁾	Готский	891	764	273	102	15	1	2046	11,24	4,13	0,82	0,77
9	Heliand ⁽³⁾	Н. Франкск.	1572	1229	452	83	14	15	3350	9,16	3,65	0,73	0,55
10	Л. Н. Толстой (Война и Мир) ⁽¹⁹⁾	Русский	466	541	391	172	64	15	1649	23,11	6,73	1,32	1,34
11	М. Ю. Лермонтов (Тамань)	Русский	469	602	404	150	55	18	1698	21,52	6,44	1,28	1,24
12	Theocritus ⁽²⁰⁾	Греческий	207	279	156	100	24	4	770	22,13	6,60	1,30	0,95

$P(1)$, $P(2)$ и $P(3)$ в зависимости от α . В качестве материала использованы тексты из различных групп языков. В верхней части графика указаны приблизительные пределы вариаций α для текстов каждой группы. График показывает, что $P(1)$, $P(2)$ и $P(3)$ почти полностью определяются свойственным данному тексту значением α . Аналогичное положение имеет место и для $P(n)$ при $n > 3$, где результаты получаются менее четкими из-за большей редкости таких длинных слов.

3. Зависимость $P(n)$ от α имеет столь всеобщий характер и соблюдается с такой неожиданной точностью, что возникает желание придать ей математическое выражение.

В качестве такого выражения может быть принят закон Пуассона:

$$P(n) = \frac{\nu^{n-1}}{(n-1)!} e^{-\nu}, \quad (2)$$

где $\nu = \alpha - 1$ (средняя многосложность).

На графике теоретические значения $P(1)$, $P(2)$ и $P(3)$ по формуле (2) нанесены пунктиром.

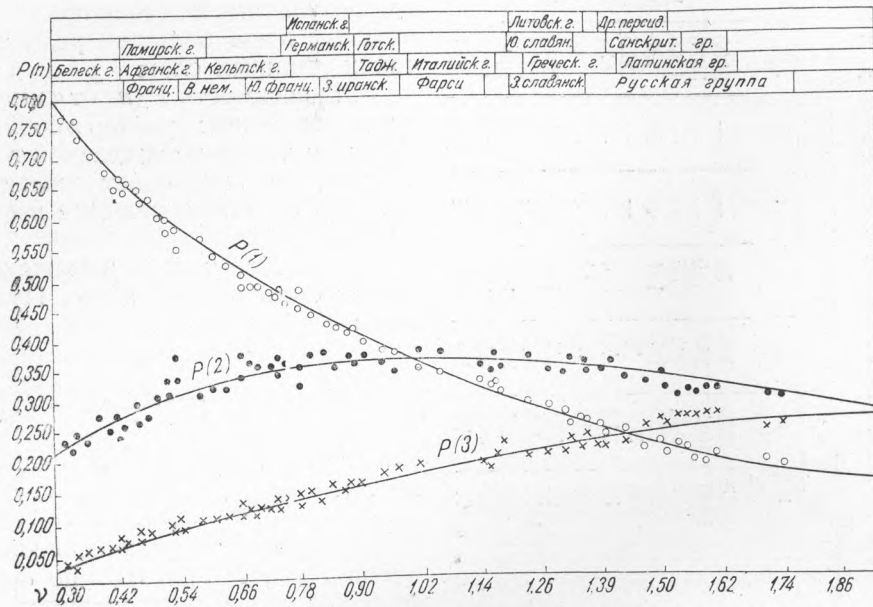
4. Дополнительной проверкой положения о подчинении укладов речи закону Пуассона является наличие согласия между старшими моментами: ν , μ_2 и μ_3 . Данные, указывающие на наличие тенденции к согласованию между тремя указанными моментами, приведены в табл. 6.

Конечно, как указал в свое время академик А. А. Марков, первый применивший стохастический метод к речевому материалу: «следует помнить, что полного совпадения чисел в подобных исследованиях, где теория соединена с опытом, нельзя требовать»⁽²¹⁾.

5. Соответственно этому и значения χ^2 и $P_k(\chi)$ (критерий согласия Пирсона) в части случаев дают хорошие данные, а в части случаев существенно отклоняются, что видно из таблицы 7.

Таблица 7

По табл. 6		Число слогов в слове (n)					χ^2	$P_k(\chi)$
		1	2	3	4	5		
№ 1	$P(n)'$	0,6250	0,2938	0,0690	0,0108	0,0013	—	—
	Nn'	1814,4	852,8	200,4	35,4	—	0,432	0,9—0,8
	Nn	1823	849	194	37	—	—	—
№ 9	$P(n)'$	0,4819	0,3518	0,1284	0,0338	0,0061	—	—
	Nn'	1613	1177	429	112	19	13,32	0,005—0,002
	Nn	1572	1229	452	83	14	—	—

График размещения выборочных данных в системе Пуассона $P(n)_\alpha$

Отклонения в значениях χ^2 и $P_k(\chi)$, подобные приведенным в табл. 7 (№ 9), повидимому, не случайны и не могут быть объяснены одной малочисленностью выборки (3350 слов).

Поступило
14 II 1946

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

- ¹ G. E. Lessing, «Laokoon», herausgeg. v. Fr. Bornmüller, B. III, Vorrede u. Abschn. I—XVI. ² Aivangelium Ulfilas, «Codex argenteus», ed. A. Uppström, Upsaliae, MDCCCLIV, Joh. c. 16, 17, Luc. c. 8. ³ Heliand, herausgeg. v. Sievers, Halle, 1878, 58—158, 159—338, 2698—2799. ⁴ Tatianus, «Diatessara», по Braune (S. 46, 47; 49—50). ⁵ Eike v. Repgow, «Sassenspiegel», W. Stammler, Hamburg, 1921. N. D. L. Buch A. ⁶ W. Geiger, Grundriss der Iranischen Philologie, I, Abt. 2, S. 287—293. ⁷ M. L'abbé Labouderic, Matériaux pour servir à l'histoire des dialectes de la langue française, Mém. d. l. Soc. R. d. Ant. de France, Paris, MDCCCXXIV, 432—545, 490. ⁸ Ibid., 492. ⁹ Ibid., 502. ¹⁰ Ibid., 503. ¹¹ Ibid., 495. ¹² Wolfram v. Eschenbach, «Parzival», herausgeg. v. K. Lachmann, Berolino, 1833, 64—114, 5; 68—124, 26. ¹³ «Nieblungen Not unt diu Klage», herausgeg. v. K. Lachmann, Berolino, 1826, (Not), S. 59—64; (Klage) v. 1—100, 4. ¹⁴ d. Goxrud (+ Kese, Zefre), B. A. Жуковский, «Материалы для изучения персидских наречий», СПб, 1888, I, стр. 8—9. ¹⁵ Grhya-samgzaha-paricishta, Gobhila-putra Z. d. D. M. G., H. IV, S. 533, Bloomfeld (1°—60°). ¹⁶ Broccans hymnen (St. Brigitta), Windisch, «Irische Texte», Leipzig, 1880 (S. 27—38). ¹⁷ Vie de St. Léger, по Chrestomathie v. Bartsch, p. 13—18. ¹⁸ Grundriss der Iranischen Philologie, Bd. I, Abt. 2 (S. 226—227), «Das Rebhuhn und der Falke». ¹⁹ Л. Н. Толстой, «Война и мир» (охота). ²⁰ Θεοφρατου Ειδύλλιον, XXV, «Ηρακλης λεοντοφονος η Αυγαιοζ κληρος». ²¹ А. А. Марков, «Исчисление вероятностей», Ленинград, 1924, 574—575.