

В.А. Карпов

ОСОБЕННОСТИ СТИЛЯ КАК СИСТЕМЫ

А для низкой жизни были числа,
как домашний, подъяремный скот,
потому что все оттенки смысла
умное число передает.

Н. Гумилев

Проверка гармонии алгеброй, а лингвистического исследования хотя бы числом, – со временем станет нормой. Тогда даже такое размытое пограничное понятие как “стиль” (лингвистика и литературоведение) предстанет как система своего рода. Для этого сегодня есть системно-философское и математическое обоснование. И этот доклад – еще одна попытка внести строгость и точность в представление о стиле как системе. При доказательстве системности стиля я воспользуюсь некоторыми количественными и фактографическими данными предыдущих работ [1; 3].

В 1995 г. при подготовке лексической базы для компьютерного учебника белорусского языка сотрудники НИЛ теоретической и прикладной лингвистики ввели в компьютер 5 стилевых частотных словарей (авторы Можейко Н.С. и Супрун А.Е.). Объем всех пяти словарей до анализа и обработки составил 105 927 слов.

Гипотетически предполагалось, что лексика стилей должна пересекаться, т. е. какая-то лексема могла входить во все 5 стилей, какая-то в 4, 3, 2 стиля. И, наконец, каждый стиль должен иметь специфическую лексику, отмеченную в относящихся только к этому стилю текстах (специфика предполагает спектр особенностей от фонетических, морфологических до собственно лексических). Сегодня можно утверждать, что эта гипотеза базируется не столько на интуиции и опыте восприятия текстов, относящихся к разным стилям (от разговорного до научного), сколько на трех законах системной философии (законы соответствия, межсистемного сходства и межсистемной симметрии) [4, с. 29–30]. Если стиль – система своего рода – значит, должно быть и межсистемное межстилевое сходство, проявляющееся через симметрию совпадения, наложения.

Вторая гипотеза касалась поэтичности белорусского языка. Автору, родившемуся “там, где резной палисад”, после знакомства с белорусским языком настойчиво казалось, что этот язык – язык поэзии.

Третья гипотеза – интуитивно-практическая – касалась того, что совсем неточно, на мой взгляд, называют словом менталитет, а я бы назвал мировосприятием белорусского народа.

Все три гипотезы нашли свое подтверждение при логико-математическом подходе.

Теоретически же при этих условиях должно существовать 32 подмножества (26 подсистем сходств и 5 подсистем специфики при одной пустой подсистеме). Сравнение содержания файлов пяти словарей программой (Я.В. Карпов) позволило выделить на указанном объеме 30 подмножеств, учитывающих различные вхождения лексем. Нет только одной подсистемы, где проза, фольклор и поэзия должны пересекаться, при этом публицистика и разговорный стиль как подсистемы должны быть пусты. После этого наложения-сопоставления объем словаря сжался почти вдвое до 55 460 единиц.

Количественные данные сведены в матрицу (см. табл. 1), где стили имеют следующие обозначения – художественная проза (Пр), публицистика (Пу), устное народное творчество (У), разговорный (Р) и поэзия (По). После плюс-минусовых кодов подсистем идут четыре столбца числовых показателей, характеризующих своеобразие каждой подсистемы с точки зрения высокой и низкой частотности. Расшифровка заголовков столбцов матрицы следующая. Столбец “общие лексем” представляет числа слов общих для всех пяти, четырех, трех, двух (суммарные данные) стилей и в конце представляет число слов специфичных для одного стиля. Второй столбец “максимальная частота” содержит данные по встречаемости наиболее употребительного слова. Третий столбец представляет предельные минимумы, допустим, $(1 + 1 + 1 + 1 + 1 = 5)$, т. е. слово отмечено по 1 разу в каждом стиле. Это хороший показатель, который увязывается со следующим столбцом “число слов с минимальной частотой” (в первой строке таких слов 5).

Таблица 1

Матрица «Дифференциально-частотный словарь пяти стилей»

	Пр	Пу	У	Р	По	Общих лексем	Макс. частота	Мин. допустим. частота	Слов с мин. частотой
1	+	+	+	+	+	3627	51800	5	5
2	–	+	+	+	+	324	1053	4	6
3	+	+	+	+	–	350	460	4	9
4	+	+	+	–	+	461	2088	4	7
5	+	–	+	+	+	1422	5904	4	30

Продолжение табл. 1

	Пр	Пу	У	Р	По	Общих лексем	Макс. частота	Мин. допустим. частота	Слов с мин. частотой
6	+	+	-	+	+	1687	4889	4	23
7	+	-	+	-	+	-	-	3	-
8	+	+	+	-	-	130	42	3	19
9	-	+	+	+	-	140	390	3	12
10	-	+	+	-	+	184	59	3	22
11	-	+	-	+	+	471	237	3	51
12	+	-	+	+	-	561	272	3	64
13	-	-	+	+	+	686	156	3	60
14	+	+	-	+	-	760	429	3	63
15	+	-	-	+	+	934	713	3	111
16	+	+	-	-	+	1033	177	3	89
17	-	+	+	-	-	182	231	2	77
18	+	-	+	-	-	543	61	2	172
19	-	+	-	+	-	766	157	2	221
20	+	-	-	+	-	905	105	2	480
21	-	-	-	+	+	952	164	2	363
22	-	+	-	-	+	1031	143	2	290
23	+	+	-	-	-	1037	340	2	301
24	-	-	+	-	+	1073	91	2	302
25	-	-	+	+	-	1248	93	2	316
26	+	-	-	-	+	1824	40	2	604
27	-	-	-	+	-	5164	253	1	3934
28	+	-	-	-	-	5434	81	1	4264
29	-	+	-	-	-	5864	300	1	3476
30	-	-	-	-	+	7892	264	1	5920
31	-	-	+	-	-	8865	142	1	5507
32	-	-	-	-	-	-	-	-	-

Подсистема 1 зона общности всех пяти стилей – это зона стабилитета, предположительно она должна содержать: а) служебные слова – предлоги, частицы, союзы – эти своеобразные языковые “шайбочки”, “винтики”, “гаечки” и пр. без которых в текстах любого стиля нам не собрать словосочетания или предложения как единицы более высокого уровня; б) местоимения – т. к. в любом стиле есть говорящий... в) многозначные слова и слова с наиболее абстрактной, нейтральной семантикой, что позволяет их использование в конкретном стиле в одном из значений. Анализ лишь первой десятки самых частых слов в каждом отдельном стиле убедительно доказывает эту гипотезу. Эти данные сведены в таблицу 2, где произведено частеречное выравнивание, нарушающее ранговую частоту, но показывающее даже без частот специфику внутри общности.

Частеречный состав первой десятки слов зоны общности

№	Проза	Публ.	Устн. нар. тв.	Разгов.	Поэзия
1	ён	ён	ён	ён	ён
2	і	і	і	і	і
3	у (ў)	у (ў)	у (ў)	у (ў)	у (ў)
4	не	не	не	не	не
5	на	на	на	на	на
6	я	гэты (гэта)	я (мы)	я (мы)	я (мы)
7	з	з	з	ты (вы)	з
8	быць	быць	ты (вы)	быць	ты (вы)
9	а	які	а	а	а
10	што (злуч.)	увесь	да	гэты (гэта)	як

Косвенно о подсистеме 1 можно судить по данным “Частотного словаря русского языка” объемом около 40 000 слов, полученных на текстах объемом свыше одного миллиона словоупотреблений. Там первые 10 слов (0,025 % всего словника) покрывают 192 777 словоупотреблений (18,25 % всех словоупотреблений), первые 100 слов (0,254 % словника) покрывают 436 949 словоупотреблений (41,36 % всех словоупотреблений), первая тысяча слов (2,54 % словника) покрывает 712 552 словоупотреблений (67,46 % всех словоупотреблений) [5, с. 924]. Как видим, 3 627 слов от словаря мощностью в 55 460 единиц составляют менее 10 % и это ядро должно перекрывать уже значительно больший, чем 67 % словоупотреблений в текстах разных стилей.

Что же мы находим в этом ядре среди автосемантических слов? Ограничимся 10–15 примерами каждой части речи, следующими по убыванию частоты:

1. Существительные : *год* 3366, *дзень* 2836, *людзі* 2790, *рука* 2437, *час* 2400, *зямля* 2254, *пан* 2288, *чалавек* 2206, *раз* 2133, *вока* 1875, *хата* 1820, *лес* 1814, *дзіця* 1728, *поле* 1661, *сэрца* 1498.

2. Глаголы : *быць* 17301, *ісці* 3637, *пайсці* 3289, *сказаць* 3203, *казаць* 3102, *магчы* 2850, *ведаць* 2677, *даць* 2469, *жыць* 2314, *хацець* 2282, *стаць* 2169, *гаварыць* 2018, *бачыць* 1980, *прыйсці* 1830, *хадзіць* 1774.

3. Прилагательные : *новы* 1756, *кожны* 1706, *добры* 1702, *малады* 1586, *вялікі* 1520, *іншы* 1282, *стары* 1127, *белы* 987, *савецкі* 931, *родны* 920, *злёны* 744, *чорны* 695, *чырвоны* 684, *цэлы* 633, *малы* 632, *беларускі* 604.

4. Наречия : *цяпер* 2640, *тады* 1811, *добра* 1521, *зноў* 1405, *вельмі* 1127, *сёння* 1008, *зусім* 771, *многа* 738, *разам* 673, *зараз* 627, *ніколі* 590, *проста* 578, *рана* 577, *сюды* 575, *далей* 573.

5. Числительные : *адзін* 5149, *два* 1905, *першы* 1832, *тысяча* 1591, *тры* 1375, *дваццаць* 819, *пяць* 755, *трэці* 643, *чатыры* 578, *сорак* 439, *дзесяць* 388, *трыццаць* 371, *семдзсят* 303, *шэсць* 286, *мільён* 278.

Какой простор открывают эти данные для построения точной и строгой концептуальной картины “БЕЛАРУСЫ” даже по отрывочным данным пяти стилей: это – *люди, хорошо живущие сегодня во времени-пространстве (на своей земле), где находятся их лес-поле-хата. Они – добрыя-великие, молодые и старые, они все видят, знают и отдают свое сердце детям* – очень значащие слова... И это еще одно новое фундаментальное направление дальнейшего исследования материала стилового дифференциального словаря (начальная гипотеза 3).

Подсистемы 2–6 представляют первую ослабленную зону варьетета, который на количественном уровне уже есть в зоне общности, где количество переходит в качество. Подсистемы 7–16 и 17–26 представляют также еще более ослабленные подзоны варьетета, переходя в подсистемах 27–31 в зону специфики.

Суммарно специфика составляет 33 219 слов. Это более половины от мощности всего словаря. И это значимо, т.к. внутри стиля есть своеобразные собственные “ядра стилей”, а слова с частотой 1–5 представляют их собственную периферию. Представлю предельно кратко наполнение подсистем специфики конкретными примерами.

Специфика художественной прозы составляет 5 434 слова (*належыць* 27, *ух!* 16, *перапытаць* 15, *парушыць* 13, *расчалка* 12, *начытаба* 12, *выйсце* 12, “*час ад часу*” 11, *разглядваючы* 11, *падміргнуць* 11, *няўпэўнена* 11, *даверца* 11).

Специфика публицистики представлена 5 864 словами (*КПСС* 300, *эканамічны* 221, *БССР* 123, *восемсот* 102, *кангрэс* 96, *імперыялізм* 92, *пяцідзсяцігоддзе* 80, *масавы* 79, *укараненне* 78, *ААН* 77, *прагрэсіўны* 71, *міралюбівы* 71, *правядзенне* 70, *фонд* 63, *забеспячэнне* 63).

Специфика устного народного творчества составляет 8 865 слов (*каляда* 142, *люлі-люлі-люлі* 94, *дай* 81, *люлі-люлі* 77, *выбіраці* 68, *пападдзя* 65, *піці* 63, *а-а* 63, *біці* 62, *сячы* 62, *гулец* 54, *шуба* 47, *анно ‘толькі’* 44, *пісар* 43, *жаці* 43).

Специфика разговорной речи представлена 5 164 словами (*стары* 253, *галаўны* 86, *бывала* 83, *неўміручасць* 45, *алі* 26, *дзесяты* 24, *усяроўна* 16, *рада* 16, *маёўка* 16, *папрадаваць* 15, *э-э-э* 14, *перагіб* 12).

Специфика поэзии составляет 7 892 слова (*нямака* 70, *мёртвы* 51, *жаўрук* 49, *гоні* 42, *днём* 40, *поўня* 38, *любая* 37, *высь* 31, *каханая* 27, *світалыны* 25, *прарасці* 25, *грымота* 25, *прамяністы* 24, *строма* 23, *спльць* 21, *мінулае* 19, *многія* 18, *гаючы* 18, *увысь* 17, *сяжына* 17, *пупышка* 16, *навальнічны* 16, *лістота* 16, *замець* 16, *спачын* 15, *пуціна* 15). Видно, что специфика каждого из стилей связана со всем комплексом языковых явлений. Так, аббревиатуры часты в публицистике, но там практически невозможны междометия, характерные для художественной прозы, устного народного творчества и разговорного стиля. Произносительные варианты являются особенностью разговорной речи, а глаголы двусложные вместо односложных (ср. *біці* и *біць*, *жаці* и *жаць* и т. п.) отмечены в устном народном творчестве, связанном с ритмом и рифмой.

Проверку на то, что выделение стилей и межстилевой лексики имеет под собой строгую математическую основу, легко сделать построив вероятностную модель явления.

С точки зрения теории вероятности априорно количественное наполнение всех подсистем мы можем считать только равновероятным при равных объемах текстов выборки. При 32 подсистемах мы имеем 6 макросостояний, которые могут иметь микросостояния. Подсистемы 1 и 32 имеют макросостояния равные микросостояниям. Макросостояния 2 и 5 имеют по 5 микросостояний (подсистемы 2–6 и 27–31), а макросостояния 3 и 4 включают по 10 микросостояний (подсистемы 7–16 и 17–26). Шестое макросостояние и микросостояния у нас пусты – их наполнения равны нулю. Исключив его из расчетов, получаем вероятность 1/31 для любого микросостояния. В числовом выражении 1/31 от объема сжатого словаря пяти стилей в 55 460 слов будет составлять 1 789 слов. С вероятностной моделью детально можно ознакомиться в указанных предыдущих работах [см. 1 и 3]. Здесь мы приведем лишь данные о количественном наполнении модели (табл. 3).

Таблица 3

Данные вероятности и реальности

1/31	5/31	10/31	10/31	5/31		их вероятные доли
1789	8945	17890	17890	8945		вероятно-числовые доли
3627	4244	4899	9561	33219		их реальные данные
+ 2.02	– 0.47	– 0.27	– 0.53	+ 3.71		превышение/уменьшение

Как видим из данных, теоретическая вероятность и реальное наполнение подсистем вступают в противоречия. Чтобы разобраться в них, построим систему противоречий “вероятность : реальность” в виде двухмерного куба, где В означает вероятные данные, а Р – реальные данные, “+” означает равное и больше, а “–” означает неравное и меньше.

Соотношение вероятностных и реальных данных в подсистемах оказалось следующим. Подсистема 1 должна представлять данные, где вероятность события соответствует реальности, что указывает на равновесность данной системы, ее строгую детерминированность. В нашем случае эта подсистема пуста. Значит, лексическая система языка является **неравновесной** (должна состоять из различного рода подсистем), что мы и имеем в реальности. А этот факт **однозначно** указывает на динамику системы. Подсистема 2 представляет данные, где вероятность превышает реальные данные. У нас это макросостояния 2, 3 и 4. Подсистема 3 представляет превышение реальных данных над вероятностными. У нас это макросостояния 1 и 5. Подсистема 4 представляет отрицательное согласие вероятности события и реальности. У нас это подсистема 32, где не ожидалось слов, которые бы не относились ни к одному стилю, т. к. любое новое слово, входящее в систему языка, должно автоматически попадать в одну из подсистем. Так, любой новый термин, допустим, физический, сначала попадает именно в свою подсистему физической терминологии, или, если шире, в подсистему научной лексики. Далее он может входить в другие терминологии, становясь более общим, и постепенно подниматься вверх, входя в другие подсистемы, что означает его уход из подсистемы специфики.

Плюсовые макросостояния-подсистемы 1 и 5 доказывают три фундаментальных лингвистических положения, характеризующих любое явление:

- 1) существование в языке-системе “ядра” лексики (подсистема 1), которое не имеет оппозиции, а, значит, является нейтральным по отношению к любому стилю (макросостояние 1). Объем этой подсистемы в сравнении с другими не столь велик (3 627), но это количество в 2 раза превышает вероятность, которую мы до опыта могли предполагать только как равную;
- 2) существование в языке-системе обширной переходной зоны, состоящей из 3-х подзон варьетета (подсистемы 7–26), где видны три разнонаправленные тенденции: стремление к общности (2–6), стремление к устойчивости (7–16) и стремление к унификации (17–26);
- 3) существование в системе языка специфической лексики, создающей

непересекающиеся подмножества, или собственную лексику конкретных языковых стилей. Превышение над вероятностью почти в 4 раза (3,71), т. е. почти в 2 раза больше, чем у “ядра” лексики.

Макросостояния 2, 3, 4, показывающие различные степени сходств подсистем, иллюстрируют следующую динамику:

а) подсистемы с 4-мя плюсами и 1-им минусом (сходств больше, чем различий) стремятся к вероятностным данным, что означает факт поддержки ядра периферией и указывает на взаимосвязи и взаимодействия в рамках целостной системы лексики;

б) подмножества с 3-мя плюсами и 2-мя минусами, где сходств также пока больше, чем различий, уже меньше поддерживают ядро системы, не случайно среди подсистем этого макросостояния уже есть пустая подсистема - микросостояние 7;

в) подмножества с 2-мя плюсами и 3-мя минусами демонстрируют переход к подсистемам специфики уменьшением реальных данных над вероятием, показывая, что сходство (пересечения) начинают вырождаться в специфику (непересечения).

Вывод: с позиций теории вероятности на уровне лексики доказаны:

1) правомерность интуитивного выделения в языке-системе пяти указанных стилей;

2) количественные неравенства, которые при анализе конкретной лексики обнаруживают качественные отличия в виде новых количественных преобладаний типа оппозиций “свое : заимствованное”, “одни морфемы : другие морфемы”, “одна фонограмма лексемы : другая фонограмма лексемы” в терминах гласных и согласных и т. п.;

3) связность подсистем, динамика их развития и переходов из одной в другую.

Гипотеза 2 о поэтичности белорусского языка доказывается следующими фактами, представленными в таблице 4.

Таблица 4

Соотношение объема отдельного словаря и специфики

	Проза	Публицистика	Устное нар.тв-во	Разговорный	Поэзия	
1	21754	18319	21000	20351	24629	мощность
2	5434	5864	8865	5164	7892	специфика
3	1789	1789	1789	1789	1789	вероятие
4	16320	11455	12135	15184	16737	пересеч-я

О чем говорят эти сухие числа? Данные второй строки показывают, что в основе белорусского языка лежат устное народное творчество и ... поэзия, превышающие вероятность почти в 4 раза, тогда как специфика остальных превышает в 3 раза.

При втором месте по объему специфики поэзия отдала в общестилевую «копилку» больше всего, что видно из четвертой строки табл. 3.

Тот же результат виден при пересечении двух стилей: поэзия – проза 1 место (есть связь с жизнью), поэзия – устное народное творчество 3 место (близость к истокам!), поэзия – публицистика 5 место (гражданская лирика) из 10 возможных. Пересечения поэзии с двумя иными стилями показывает, что опять поэзия в 3-х случаях из 10 впереди: поэзия-проза-публицистика (1 место), поэзия-проза-разговорный стиль (2 место), поэзия-разговорный стиль-устное народное творчество (4 место). И по пересечениям четырех стилей 3 “медальных” места и одно призовое из 5 достаются поэзии: поэзия-разговорный-публицистика-проза (1 место), поэзия-разговорный-проза-устное народное творчество 2, поэзия-публицистика-проза-устное народное творчество (3).

Как видите, число позволило нам доказать маркированность, внутреннюю неявную поэтичность белорусского языка по первым местам, где обязательно оказывается поэзия.

Вместо сухого заключения, где опять надо в деталях показывать, что инструментарий в виде матриц-классификаторов и матрицы вероятности несложны и могут использоваться в любых работах по языку, я закончу стихами:

Число бесстрастно, но какую страсть
мы видим в жизни, равной единице,
делимой на мгновения, страницы...,
где только слову отдается власть.

Литература

1. Карпов В.А. Дифференциально-частотный стилевой словарь белорусского языка // Письменник – мова – стыль. Тэз. дакл. і павед. міжнар. навук. канф., прысв. 70-годдзюпраф. Л.М. Шакуна. Мінск, 1996.
2. Карпов В.А. Текст о тексте в тексте. Язык и социум // Матер. VI Междун. научн. конф. Ч. II. Мінск: 2004. С. 40–48.
3. Карпаў Уладзімір. Слоўнік пяці стыляў беларускай мовы // Роднае слова. № 9. С. 51–54.
4. Урманцев Ю.А. Девять плюс один этюд о системной философии. М.: Современные тетради, 2001.
5. Частотный словарь русского языка. Под редакцией Л.Н. Засориной. М.: Русский язык, 1977.