

ИЗ ОПЫТА СОСТАВЛЕНИЯ ЧАСТОТНОГО СЛОВАРЯ ЯЗЫКА К. ПАУСТОВСКОГО

Л. СУДАВИЧЕНЕ

Настоящая статья представляет собою сообщение о проводимых исследованиях на кафедре русского языка Вильнюсского государственного университета им. В. Капукаса¹. В статье не ставится задача осветить лексикографические и лексико-статистические принципы составления словаря писателя², в ней лишь представлены некоторые данные, выявленные в результате статистического анализа романа К. Паустовского „Дым отечества“³. Статистический анализ этого произведения является частью большой работы — подготовки Частотного словаря языка К. Паустовского и Словаря языка писателя⁴.

Роман „Дым отечества“ был написан К. Паустовским в 1944 г., примерно за год до окончания Великой Отечественной войны. Рукопись его была потеряна в том же 1944 г. и случайно найдена лишь в 1963 г. Автор опубликовал ее в „нетронутым виде“, — как пишет он сам, — „как одно из небольших свидетельств эпохи“. Писатель ничего не изменил в романе, так как считал, что „невозможно и ненужно подгонять каждую вещь, написанную давно, к требованиям сегодняшнего дня и тем самым снимать с нее признаки времени“⁵. Таким образом, было выбрано для распски издание, проверенное и одобренное самим автором — К. Паустовским, писателем, „влюбленным

¹ Информацию о составлении Словаря языка К. Паустовского см.: „Уч. зап. высш. зав. Лит. ССР“. Языкознание, т. ЖХШ. Вильнюс, 1970.

² Это будет сделано в специальной работе. Разработкой отдельных вопросов занимаются студенты-русисты. См.: Ж. Ужжурелите. Построение словарной статьи в словаре языка К. Паустовского. — Краткое содержание докладов III межвуз. студ. научн.-филолог. конф. Ленинград. госуд. пед. инст. им. А. И. Герцена. Л., 1970, стр. 30.

³ К. Г. Паустовский. Дым отечества. М., Изд. „Сов. Россия“, 1964.

⁴ В работе принимают участие преподаватели кафедры русского языка ВГУ: Э. Галнайтите, З. Д. аунене, Т. Плакунова, Л. Судавичене и др.; студенты-русисты, участники спецсеминара „Словарь языка писателя“: Н. Воробьева, Т. Шкурко, И. Волкова, Р. Пекелите, Ж. Ужжурелите, В. Павлова, С. Соколова, В. Усенко и др.

⁵ К. Г. Паустовский. Дым отечества, М., изд. „Сов. Россия“, 1964, „От автора“.

в русское слово", человеком, „у которого чувство прекрасного было развито до совершенства“⁶.

В результате сплошной расписки текста была создана картотека, объемом около 57.000 словоформ⁷. Полученная таким образом картотека составлена по алфавиту. Материалы этой картотеки были использованы и для подготовки Частотного словаря указанного романа. Для этой цели были подсчитаны все словоформы и выделены для каждого слова исходные формы (репрезентант слова⁸), которые явились единицами Частотного списка. Слова в Частотном списке расположены по убывающей частоте. Первым в таком списке оказывается слово с наивысшей частотой. В нашем Словаре (списке) таким словом оказался предлог *и* и его вариант *во* (см. таблицу 3).

Статистическая структура словаря представлена в таблице 1. В первой графе ее указана частота слова (*F*), во второй — количество слов, имеющих данную частоту (*m*). Таблица 1 дает возможность установить, сколько имеется в романе слов, а также сколько слов встретилось не более *n* раз. Из таблицы видно, что с уменьшением частоты слов объем группы слов с данной частотой закономерно увеличивается. Так, число слов с частотой 1 составляет почти половину списка (3.798 из 7.827).

Таблица 1

F	m	F	m	F	m	F	m	F	m	F	m
1920	1	387	1	274	1	142	1	120	1	97	1
1875	1	376	1	271	1	141	1	119	2	95	1
1286	1	372	1	208	1	140	1	117	1	94	1
1036	1	360	1	201	1	136	2	116	1	91	2
914	1	351	1	190	2	134	1	112	1	90	1
787	1	347	1	184	1	133	1	109	1	88	2
758	1	340	1	179	1	132	1	108	4	87	1
618	1	335	1	165	1	130	1	106	2	86	3
606	1	312	1	157	2	128	1	104	2	85	2
544	1	307	1	156	2	127	1	103	1	84	2
516	1	292	1	153	1	125	1	100	1	83	3
393	1	278	1	152	2	121	1	99	1	82	4

⁶ Газета „Неделя“, № 30 (438); № 33 (441), 1968.

⁷ Частота имен собственных не учитывалась. Так, например, фамилия Пахомов встретилась в тексте около 60 раз, а в словаре она включена в число слов с частотой 1.

⁸ Термин взят из работы П. М. Алексеева „Частотные словари и приемы их составления“. — В кн. „Статистика речи“, Л., изд. „Наука“, Ленинградское отделение, 1968.

F	m	F	m	F	m	F	m	F	m	F	m
81	1	63	2	48	1	34	8	20	18	6	209
80	1	62	3	47	2	33	12	19	23	5	264
78	1	61	2	46	7	32	9	18	19	4	393
77	1	60	1	45	8	31	5	17	36	3	705
76	1	58	1	44	4	30	16	16	25	2	1277
74	3	57	2	43	9	29	10	15	33	1	3798
73	6	56	2	42	4	28	7	14	36		
72	1	55	3	41	4	27	11	13	60		
71	2	54	3	40	1	26	11	12	59		
69	3	53	1	39	6	25	12	11	58		
68	4	52	3	38	1	24	17	10	84		
67	2	51	3	37	7	23	13	9	87		
66	4	50	3	36	7	22	18	8	92		
65	1	49	6	35	4	21	23	7	153		

Обследованный текст содержит 7.827 слов, а его длина равна 57.000 словоформам. Материалом для некоторых наблюдений послужила группа самых частых слов (см. таблицу 3), предельной частотой которой является 8. В данную группу (список) вошло 1027 слов, что составляет 42.858 словоупотреблений. Это количество слов занимает около 76% всего текста — романа „Дым отечества“.

Таблица 2 дает представление об относительной употребительности некоторых частей речи.

Таблица 2

Часть речи	Количество слов	%
Имя существительное	375	36,5
Глагол	285	27,5
Имя прилагательное	133	13
Наречие	111	10,8
Предлог	31	3
Союз	17	1,6

Сопоставляя полученные сведения с данными, приводимыми Э. А. Штейнфельдт для современного русского литературного языка⁹, можно заметить,

⁹ Э. А. Штейнфельдт. Частотный словарь современного русского литературного языка, Таллин, 1963, стр. 40.

что несмотря на то, что выборка у нас была намного меньше, чем в Словаре Э. Штейнфельдт, соотношение частей речи является одинаковым. В анализируемом нами тексте так же, как и в указанном Словаре, на первом месте выступают имена существительные, немного отстает от них глагол, а затем уже идут имя прилагательное и наречие.

Сравнивая группы слов с определенной частотой, можно обнаружить, что соотношения различных частей речи внутри их являются неодинаковыми. Так, в группе 100 самых частых слов предлоги, союзы и частицы составляют 36%, существительные — 19%, глагол — 18%, прилагательное — 1%, числительное — 1%. По мере увеличения словаря число знаменательных частей речи возрастает, а число служебных частей речи сокращается¹⁰.

Анализ грамматических классов слов позволил отметить большое разнообразие слов как в лексико-тематическом, так и в грамматическом плане. Так, среди частых имен существительных довольно много военных терминов, что объясняется характером произведения. Ср.: блиндаж, боец, бой, взрыв, война, землянка, зенитка, капитан, осколок, санитар, солдат, снаряд, танк, шинель и др. Значительную часть составляют слова с конкретным значением: глаз, голова, день, дом, земля, комната, лес, лицо, нога, окно, пальто, письмо, плечо, пол, рука и др. Немало слов, обозначающих родство: бабушка, брат, дочь, жена, мать, муж, отец, сестра, тетя и др. Среди глаголов особенно много переходных, например, сказать, видеть, писать, найти, читать, принести, вспомнить, показать, открыть и т.п., встречаются видовые пары: понять — понимать, рассказать — рассказывать, говорить — сказать, курить — закурить, пить — выпить, писать — написать, дать — давать, кричать — крикнуть, брать — взять и др. Разнообразны глаголы и по лексическим значениям, ср.: войти, выйти, уйти, прийти, пройти, подойти, зайти, дойти, сойти, перейти; открыть, закрыть и т.п. В группе имен прилагательных много антонимических пар: маленький — большой, тяжелый — легкий, высокий — низкий, близкий — далекий, густой — редкий, мокрый — сухой, старый — молодой, хороший — плохой, крупный — мелкий и др. Можно выделить и тематическую группу слов, обозначающих цвет или оттенки цветов: черный, белый, серый, красный, желтый, синий, зеленый, розовый, рыжий, темный, светлый.

Все эти данные, по нашему мнению, представляют известный интерес и могут быть использованы в различных исследованиях и, в частности, при под-

¹⁰ Так, например, количество служебных частей речи во второй тысяче слов уменьшается в 23 раза по сравнению с первой тысячей. Анализ второй тысячи слов сделан в курсовой работе Р. Пекелите.

готовке лексических, грамматических и тематических минимумов, что особенно актуально для нашей республики¹¹.

Таблица 3

Список наиболее частых слов¹²

В (во) – 1920; и – 1875; на – 1286; не – 1036; быть – 914; он – 787; с (со) – 758; ли – 618; что (С) – 606; я – 544; она – 516; как – 393; сказать – 387; за – 376; но – 372; к (ко) – 360; что (М) – 351; это – 347; этот – 340; а – 335; у – 312; из (изо) – 307; вы – 292; от – 278; все – 274; по – 271; же (ж) – 208; то – 201; о (об) – 190, только; мочь – 184; они – 179; знать – 165; рука – 157, ты; день – 156, так; свой – 153; весь – 152, глаз; уже (уж) – 142; где – 141; спросить – 140; потом – 136, человек; до – 134; говорить – 133; такой – 132; идти – 130; дом – 128; когда – 127; еще – 125; один – 121; ответить – 120; голова – 119, да; над – 117; стоять – 116; вот – 112; себя – 109; бы – 108, вода, смотреть, чтобы; жизнь – 106, ни; дорога – 104, нет (нету); ночь – 103; мы – 100; под – 99; окно – 97; слово – 95; очень – 94; должен – 91, думать; здесь – 90; надо – 88, сейчас; казаться – 87; земля – 86, каждый, там; ветер – 85, ну; выйти – 84, даже; небо – 83, пойти, через; видеть – 82, люди, посмотреть, снег; всё – 81; начать – 80; лицо – 78; долго – 77; для – 76; машина – 74, сам, уйти; жить – 73, какой, маленький, ничто, увидеть, хотеть; дверь – 72; берег – 71, старый; портрет – 69, раз, сидеть; голос – 68, несколько, около, сестра; будто – 67, тот; город – 66, лежать, мой, черный; свет – 65; лес – 63, прийти; море – 62, пройти, хорошо; остаться – 61, стена; сесть – 60; всегда – 58; стать – 57, теперь; женщина – 56, солнце; подумать – 55, тогда, отец; если – 54, после, рядом; капитан – 53; вечер – 52, два, её; девушка – 51, кто, слышать; вернуться – 50, взять, понимать; время – 49, или, любить, письмо, снова, театр; понять – 48; вещь – 47, воздух; его – 46, никто, сад, совсем, старик, стол, час; белый – 45, война, много, наш, обновиться, огонь, темнота, что-то; друг – 44, конец, первый, плечо; гора – 43, дождь, комната, место, молчать, поезд, последний, самый, слушать; без – 42, конечно, нога, рассказать; никогда – 41, писать, приехать, тот-час; сердце – 40; большой – 39, войти, далеко, другой, почему, цвет (цветок); наконец – 38; который – 37, перед, похожий, тоже, тяжелый, улица,

¹¹ См. „Успехи, свершения, поиски“ – Журн. „Русский язык в национальной школе“, 1967, № 5, стр. 18–19.

¹² Рядом со словом указывается частота. Если частота повторяется, то она отмечается только рядом с первым словом. В скобках дается вариант слова или указывается часть речи: С – союз, М – местоимение, Суш – существительное.

чай; волос — 36, ехать, запах, книга, найти, написать, утро; год — 35, пальто, стекло, тихо; боец — 34, лист, можно, нельзя, поднять, почти, уходить, художник; вместе — 33, ждать, краска, куда, неужели, открыть, пол, решить, самолет, трава, три, холодный; вздохнуть — 32, встать, высокий, делать, дым, зима, милый, новый, среди; закрыть — 31, пыль, сделать, туман, узнать; бог — 30, вспомнить, дело, железный, из-за, кричать, любовь, минута, мысль, пахнуть, работать, скоро, спать, страна, сухой, шофер; начаться — 29, ничего, огромный, подойти, пора, принести, сила, счастье, темный, часто; дать — 28, карман, медленно, потому, пустой, случиться, тут; взглянуть — 27, враг, висеть, глядеть, кто-то, мокрый, откуда, разговор, серый, сторона, хочется; бояться — 26, бежать, одна, очевидно, рассвет, снять, стараться, стих, улыбаться, умереть, чем; вниз — 25, внизу, знакомый, иногда, крикнуть, кроме, немец, подняться, пока, при, теплый, ящик; гореть — 24, давно, зачем, звезда, каюта, квартира, кресло, месяц, начинать, платок, пушкинский, работа, река, русский, утром, читать, шинель; второй — 23, диван, есть, зал, заметить, звать, мальчик, опять, поехать, приходить, разговаривать, считать, уехать; великий — 22, гостиница, дерево, значит, из-под, капля, красный, мост, оттуда, палец, палуба, показать, поле, помочь, сон, услышать, хотя, чужой; блиндаж — 21, больше, вагон, волна, держать, достать, желтый, завтра, изредка, крепко, мама, молодой, морской, начинаться, низкий, нужно, осторожно, овраг, пароход, петь, положить, спина, сразу; вокзал — 20, вокруг, гудеть, длинный, звук, играть, какой-то, краснофлотец, облако, ощущение, оно, платье, синий, сквозь, стук, ходить, шуметь, язык; блеск — 19, ваш, вспоминать, девочка, закричать, колено, лампа, немецкий, очки, показаться, попасть, появиться, проснуться, просто, спрашивать, спуститься, телеграмма, труд, трудно, упасть, хороший, цвет, часы; выходить — 18, выпить, горячий, засмеяться, изба, камень, качаться, коса, лыжник, мать, открытый, подыматься, полный, пробормотать, радость, согласиться, туча, тысяча, шаг; блестящий — 17, бой, бывать, быстро, взрыв, военный (Суц), возвращаться, дача, дети, единственный, койка, лечь, лампочка, мало, останавливаться, падать, папироса, печка, писатель, пить, плакать, попросить, правый, сегодня, слеза, случай, смерть, смеяться, спокойно, стакан, стул, тишина, угол, улыбнуться, черт, щека; борт — 16, буря, век, весна, давать, живой, забыть, замолчать, зимний, край, куст, мальчишка, между, недавно, незнакомый, немного, оглянуться, парк, снаряд, совершенно, станция, сумерки, сюда, третий, шум; болеть — 15, ветка, впервые, горный, губа (губы), действительно, детство, дойти, зеленый, имя, испанский, их, красавица, ладонь, лестница, мотор, настоящий, оставаться, отвечать, позвать, помнить, пригодиться, происходить, прочесть, раз-

ный, седой, соседний, странный, стучать, тонкий, тянуться, уснуть, шоссе; благодарить — 14, бросить, валяться, верить, веселый, встретить, встреча, глубокий, греметь, дуть, зайти, километр, курить, мир, могила, нести, нужный, обернуться, оставить, пожалуй, плохо, поставить, поэт, путь, рукав, ручей, сильно, собор, сосна, спускаться, степь, туда, умирать, фонарь, чувствовать, юг; балка — 13, больной, ведь, вести, вино, всякий, грудь, густой, далекий, деревня, задуматься, зажечь, закурить, заросль, иллюминатор, лейтенант, лететь, мимо, многих, муха, назад, никакой, одеяло, однажды, опустить, пар, перестать, пожать, пожилой, позвонить, потерять, прижать, проходить, прямо, пусть, разве, рассказывать, санаторий, сапог, сделать, север, след, смочь, собака, столик, столовая, страшно, твердый, темнеть, тень, тепло, товарищ, торопливо, труба, усмехнуться, успокоиться, ухо, хоть, школа, шляпа; актер — 12, бархатный, бить, бледный, боль, вдаль, вслед, глухой, грязь, двигаться, деревянный, десять, дно, дочь, жена, землянка, иначе, кабина, каменный, красота, крутой, ледяной, мертвый; музей, напоминать, неожиданно, никак, обнять, оказаться, освещенный, остров, отъезд, поблескивать, покраснеть, порт, потащить, поцеловать, почувствовать, привыкнуть, простой, протянуть, рама, рассказ, рыться, садиться, следующий, сознание, старинный, страшный, судьба, тетя, худой, хозяйка, церковь, чайник, человеческий, чемодан, чистый, широкий; болезнь — 11, бормотать, бумага, волнение, вскоре, вскочить, газета, горизонт, грохот, гудок, дежурный (Суш), жест, заговорить, замечать, запад, звон, зеркало, коридор, крыша, кофе, легкий, легко, летá, лоб, мезонин, моряк, ночной, песок, помолчать, приезжать, просить, прошлое (Суш), птица, пыльный, рад, раньше, рассердиться, режиссер, розовый, сверкать, светлый, сердитый, слишком, спасти, становиться, старушка, строгий, схватить, танк, толстый, трамвай, туфля, ударить, удивиться, ум, ход, холм, чувство; брат — 10, вдали, весело, громко, грузовик, гул, даль, дорогóй, дрожать, душа, дыхание, затихнуть, золотой, известный, искать, испуганно, история, какой-нибудь, катер, качать, кормá, крик, крупный, кухня, лед, лето, литься, луч, любимый, метель, множество, муж, мусор, мчаться, мягкий, навстречу, нравиться, озеро, освещать, осколок, открывать, отсюда, поворот, повторить, послать, прекрасный, приезд, прислушиваться, провести, проводить, прогнить, прохладный, пустить, равно, рассматривать, ресторан, родной, роща, рояль, свеча, свист, свистеть, сено, синева, сирена, скатерть, сколько, смех, сначала, снимать, создать, солдат, спокойный, спрятать, старуха, существовать, тревога, уверенный, успеть, устроить, фреска, честный, чуть, электрический; бабушка — 9, благодарность, близкий, близко, будущий, буква, буфет, вверх, вдруг, вздрагивать, вместо, внимание, волноваться, все-таки, входить, глу-

бина, гром, готовый, доска, дрова, женский, запеть, заставить, застенчивый, затихать, звенеть, значить, иметь, искусство, как-то, командир, кто-нибудь, куртка, ложиться, матушка, мелкий, молоко, мохнатый, набережная, наверное, накинуть, наклониться, народ, нестись, отвезти, ответ, отдых, палата, память, пес, пешком, площадка, побежать, поблагодарить, погодить, ползти, поэзия, правда, представить, привезти, против, разноцветный, редкий, решиться, родина, санитар, свежий, светить, свободный, сей, серьезный, снежный, собираться, сойти, стеной, строго, счастливый, сырой, трещать, тусклый, тьма, устать, холодно, целый, что-нибудь, этаж, яркий; беседка – 8, бесконечный, беспокоиться, брат, бродить, броситься, болтливый, везти, вечером, вздрогнуть, вздыхать, вид, влажный, внезапно, водитель, вопрос, воспоминание, вполголоса, встретиться, дверца, двор, добавить, догадываться, дикий, едва, зенитка, испуганный, испытывать, куча, косматый, левый, мастер, молчаливый, навсегда, норвежец, носилки, облегчение, одесский, ожидать, особенно, отправить, передать, перейти, плотный, плохой, подоконник, подходить, подымать, покрыть, половина, помогать, попрощаться, постель, прибрать, прикрыть, прислушаться, про, проволока, протереть, пятно, пять, рабочий (Суш), равнина, расти, редко, ресница, роза, рожь, светиться, сжать, скала, скамейка, собирать, солнечный, стебель, стоять, столб, столетний, телефон, тем, трубка, туманный, удар, удивляться, улыбка, ущелье, часть, чудесный, шепот, шестой, шея, ясно.

Вильнюсский государственный университет
им. В. Капсукаса

Сентябрь, 1969

DĖL K. PAUSTOVSKIO KALBOS DAŽNU MINIO ŽODYNO SUDARYMO

L. SUDAVIČIENĖ

Reziumė

Vilniaus Valstybinio V. Kapsuko Universiteto rusų kalbos katedroje pradėtas didelis kolektyvinis darbas – sudaromas K. Paustovskio dažnuminis žodynas ir autoriaus kalbos žodynas.

Štraipsnyje yra pateikta K. Paustovskio romano „Дым отечества“ statistinė struktūra. Nustatyta, kad 1000 dažnų žodžių sudaro 76% viso teksto. Nustatyti taip pat atskirų kalbos dalių kartojimosi dėsningumai. Pvz., pirmoje vietoje yra daiktavardis, antroje – veiksmažodis. Mėnėti duomenys patvirtina bendrus kalbos dalių santykiavimo dėsningumus, analizuojant grožinės literatūros tekstus (žr. G. Joselson, E. Šteinfeldt).