ОПИСАНИЕ СТАТИСТИЧЕСКОЙ СТРУКТУРЫ ТЕЛЕГРАФНЫХ ТЕКСТОВ

НАЗИК МЕЛИКЯН

Материалом для описания статистической структуры телеграмм послужили количественные данные по лексике и грамматике, извлеченные из частотного микрословаря подъязыка телеграмм. Частотный микрословарь подъязыка телеграмм составлен на материале 3000 (исходящих) телеграмм[1], поданных на Кировском почтамте г. Москвы примерно с марта по июль 1963 года[2]. Общая длина исследованных текстов[3] равна 23978 словоупотреблениям.

Частотный микрословарь подъязыка телеграмм состоит из 2031 лексемы и покрывает текст длиной в 19151 словоупотребление.

В частотный микрословарь не включены встретившиеся в исследуемых текстах: 1) числительные (в любом графическом облике); 2) имена собственные — различных типов (в том числе и имена прилагательные, образованные от имен существительных собственных — ойконимов и функционирующие то как имена собственные, то как нарицательные) На долю перечисленного в 1) и 2) приходится 4827 словоупотреблений. Помещенные в отдельный список, слова эти будут рассмотрены позднее.

Статистическая структура исследованных текстов[4]. 2031 лексема

---

[1] Ввиду малых размеров выборки (см. рекомендации П. М. Алексеева по составлению частотных словарей в сборнике «Статистика речи», Л., 1968, стр. 61—63) составленный нами частотный микрословарь не обладает необходимой точностью в определении частот; однако он является вполне достаточным по объему для того, чтобы с его помощью (на основании количественных характеристик, пусть весьма относительных по значению) получить представление о специфике отбора лексики, об основных тенденциях в распределении лексем и частей речи в телеграфных текстах, а также о том, однородна ли статистическая структура телеграмм. Ср.: «Минимальным отрезком текста для статистических подсчетов является текст или группа текстов общей длиной в 1000 словоупотреблений» (П. М. Алексеев, Частотные словари и приемы их составления, «Статистика речи», Л., 1968, стр. 62).

[2] Более точно: выборка А—991 телеграмма (967, поданных с 12 по 23 марта, и 24 телеграммы вразброс за январь, февраль и апрель); выборка Б—1000 телеграмм, поданных с 25 мая по 15 июня; выборка В—1000 телеграмм, поданных с 6 июня по 12 июля. Таким образом, была исключена возможность появления в выборке поздравительных телеграмм, поданных в связи с празднованием официально установленных торжественных дат.

[3] Рассматривается лишь собственно текст телеграммы (сообщение); адрес, пометы телеграфиста и подпись отправителя из рассмотрения исключаются.

[4] За недостатком места в данной публикации, частотный микрословарь телеграмм будет напечатан в сборнике «Вопросы структуры текста», Ереван, 1977 (находится в печати).

частотного микрословаря по частотности лексем (и длине покрываемого ими текста) распределяется следующим образом[5]:

Таблица 1

i	F	N	N, в%
1—41	856—100	9476	49,48
42—129	98—25	3932	20,54
130—397	24—6	3038	15,86
398—977	5—2	1651	8,62
978—2031	1	1054	5,50

Таким образом, 41 наиболее часто употребляющаяся лексема покрывает почти половину исследованного текста, в то время как 1054 лексемы с частотой 1 (составляющие более половины частотного микрословаря) — всего 5,5%.

Распределение лексем (и соответствующего им числа словоупотреблений) по частям речи показано ниже:

Таблица 2

Часть речи	С	Гл	П	Н	Пр	Со	Пч	Дч
L[6]	1022	406	329	161	26	16	74	1
N	8298	6154	1774	1739	771	278	136	1
N, в%	43,33	32,13	9,26	9,1	4,02	1,45	0,71	—

Из таблицы видно, что морфологически язык телеграмм — это язык, состоящий преимущественно из существительных и глаголов (75,46% от общего числа словоупотреблений). Служебные слова представлены минимально[7] (6,18% от общего числа словоупотреблений).

Прилагательные и наречия занимают промежуточное положение (18,36% от общего числа словоупотреблений).

---

[5] Применяемые в статье обозначения:
F — абсолютная частота встречаемости лексемы в данной выборке;
i — порядковый номер лексемы в словаре (ее ранг);
L — количество лексем в словаре (объем словника);
N (или *n*) — сумма словоупотреблений (объем выборки);
m — количество лексем, имеющих данную частоту;
С — существит.; П—прилагат., Гл—глагол, Н—наречие, Пр—предлог, Со—союз, Пч—причастие, Дч—деепричастие, Кр—краткая форма П или Пч; Срв—сравнительная форма П; Мд—признак модальности у П или Н.

[6] L в данном случае равен 2035, т. к. омографы, имеющие один и тот же порядковый номер (i) в микрословаре, оказались здесь распределенными между разными частями речи (например, «что»—Со и «что»—С).

[7] Класс Пр уменьшился бы существенно в количестве, не будь избыточного, а следовательно, ошибочного (с точки зрения правил составления телеграмм) употребления предлогов.

В выборках А, Б, В[8], представляющих отдельные (примерно равные по числу телеграмм) пласты исследуемого массива телеграфных текстов, данные по распределению лексем (и соответствующего им числа словоупотреблений) по частям речи таковы:

В выборке А (L_A—1134 л.; N_A—7005 словоупотреблений):

Таблица 3

Части речи	С	Гл	П	Н	Пр	Со	Пч	Дч
L	550	252	176	105	19	8	24	—
N	3244	2191	598	600	254	85	33	—

В выборке Б ($L_Б$—1072 л.; $N_Б$—5985 словоупотреблений):

Таблица 4

Части речи	С	Гл	П	Н	Пр	Со	Пч	Дч
L[9]	546	207	150	108	18	8	35	1
N	2695	1926	457	503	275	74	54	1

В выборке В (L_B—1077 л.; N_B— 6161 словоупотребление)

Таблица 5

Части речи	С	Гл	П	Н	Пр	Со	Пч	Дч
L[10]	508	231	171	100	23	11	36	—
N	2716	2062	421	591	231	93	47	—

Величины (в процентном выражении), показывающие долю[11] определенной части речи в каждой из выборок, оказываются близкими:

Таблица 6

Выборки	N, в %	С	Гл	П	Н	Со	Пр	Пч	Дч
А	100	46,30	31,27	8,53	8,6	1,21	3,62	0,47	—
Б	100	45,03	32,2	7,64	8,40	1,23	4,6	0,90	—
В	100	44,1	33,46	6,83	9,6	1,50	3,75	0,76	—

[8] См. сноску 2.

[9] L здесь равен 1073 (а не 1072); см. объяснение в сноске 6.

[10] L здесь равен 1080 (а не 1077); см. объяснение в сноске 6.

[11] Доля—отношение наблюдаемой частоты к длине выборки. «Формула для вычисления доли аналогична формуле вероятности: $p = \frac{m}{n}$ (в формуле вероятности принято давать большую, прописную букву Р, в формуле доли — малую, строчную)» — Б. Н. Головин, Язык и статистика, М., 1971, стр. 36—37.

Используя приемы изучения статистической структуры текста[12], попытаемся убедиться в том, что колебания долей, соответствующих определенной части речи в разных выборках, несущественны, подчинены одному и тому же статистическому закону распределения (даже при их наибольшем расхождении).

Сравним доли П (в попарно взятых выборках), затем доли Н, Пч и Пр (последовательно).

При сравнении долей применим формулу среднего квадратического отклонения доли[13]: $\varepsilon_{1,2} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$, (1)

где $\bar{p}$ — средняя (для двух совокупностей) доля изучаемого явления;

$\bar{q}$ — средняя доля прочих явлений;

n_1; n_2 — размеры выборок.

Вышеуказанная формула применяется при условии равного объема выборок (n_1 и n_2).

Поскольку А, Б и В не равны (по числу словоупотреблений), n_1 и n_2 (n_A и $n_Б$, если, к примеру, сравниваются доли в выборках А и Б) получат численное значение по формуле [14]:

$$n_1 = \frac{(P_A + P_Б)}{N_A + N_Б} N_A; \quad n_2 = \frac{(P_A + P_Б)}{N_A + N_Б} N_Б, \qquad (2)$$

где P_A — доля П в выборке А; $P_Б$ — доля П в выборке Б.

Подставив численные значения n_1 и n_2 в формулу среднего квадратического отклонения доли и проделав все необходимые подсчеты, получим:

Таблица 7

	$\varepsilon_{А,Б}$	$\varepsilon_{А,В}$	$\varepsilon_{Б,В}$
1) для П	1,5	1,47	1,49
2) для Н	1,35	1,39	1,38
3) для Пр	1,39 (1,389)	1,4	1,38 (1,378)
4) для Пч	1,41	1,41	1,41

Среднее квадратическое отклонение доли в каждом из рассмотренных случаев превышает разность соответствующих долей[15], поэтому можно допустить, что колебания рассматриваемых долей объясняются

---

[12] Там же, стр. 36—40.

[13] Формула «...позволяет уловить некоторые усредненные колебания долей около их средней теоретической величины» (там же, стр. 37).

[14] Б. Н. Головин, указ соч., стр. 34, 35 (где вычисления по формуле (2) выполнены для получения выборочной средней частоты при неравных выборках).

[15] Разность долей в выборках А и Б, А и В, Б и В соответственно: для П — 0,0089, 0,017, 0,008; для Н — 0,002, 0,010, 0,012; для Пр — 0,0098; 0,0013, 0,0085; для Пч — 0,0043,0,0029,0,0014.

варьированием одной и той же вероятности распределения частей речи в телеграфных текстах[16].

Приведем другие данные, также свидетельствующие об однородности статистической структуры телеграфных текстов в исследуемых пластах.

Отношение числа словоупотреблений в выборках А, Б и В[17] к числу словоупотреблений в совокупной выборке равно соответственно: 36,6%; 31,2%; 32,2%.

Среднее квадратическое отклонение полученных величин от теоретической средней ($\bar{x} = 33{,}3$) вычислено по формуле

$s = \sqrt{\frac{\sum a^2}{k}}$ (3) (где $\sum a^2$ сумма квадратов отклонений полученных величин от теоретической средней ($\bar{x}$), а к — число выборок)[18].

На основании полученной для среднего квадратического отклонения величины ($\sigma = 2{,}33$) вычислим коэффициент вариации, т. е. найдем меру колеблемости указанных величин относительно теоретической средней. Коэффициент вариации, вычисленный по формуле $v = \frac{100\,\sigma}{\bar{x}}$ (4), оказался равным 7%; таким образом, можно допустить[19], что распределение словоупотреблений в выборках, обслуживающих примерно равное количество телеграмм, подчиняется одной и той же статистической закономерности[20].

Определим еще, какова доля словоупотреблений, соответствующая лексемам с однократным употреблением ($F = 1$)[21], в каждой выборке.

Число лексем с однократным употреблением (m_1) составляет: в выборке А—635, в выборке Б—566, в выборке В—604. Соответственно, доля m_1 равна: в выборке А—9%, в выборке Б—9,4%, в выборке В—9,9%.

---

[16] «Если квадратичное отклонение доли меньше разности долей втрое (и более), мы вправе отвергнуть гипотезу о случайности, т. е. несущественности расхождения долей» (Б. Н. Головин, указ. соч., стр. 38).

[17] Напомним, что выборки А, Б, В обслуживают каждая примерно 1000 телеграмм (991, 1000 и 1010, соответственно). Разница в 10 телеграмм, т. е. примерно в 70 словоупотреблений, не может существенно повлиять на результаты подсчетов. Отметим лишь, что при точном соответствии указанных выборок по числу телеграмм, выборки Б и В окажутся почти равными друг другу и по числу словоупотреблений, отграничиваясь в то же время от выборки А (мартовской).

[18] Б. Н. Головин, указ. соч., стр. 22—24.

[19] Принято величину коэффициента вариации считать большой, т. е. вызывающей недоверие «к гипотезе о варьировании, если он превосходит 40%» (Б. Н. Головин, указ. соч., стр. 35).

[20] Задача исследования статистической структуры телеграфных текстов на широком материале — установить, зависят ли колебания величин (о которых шла речь выше) от различных факторов, таких, например, как время (месяц, сезон года), в течение которого были поданы телеграммы, вошедшие в сопоставляемые выборки.

[21] В этом случае число лексем и число словоупотреблений оказывается равным.

Нетрудно убедиться с помощью уже известной формулы среднего квадратического отклонения доли (1), $\varepsilon_{1,2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, что и в этом случае полученные величины не нарушают пределов, допустимых для случайных колебаний долей.

В попарно взятых выборках (А и Б, А и В, Б и В) число лексем с однократным употреблением составило: для совокупности N_A и $N_Б$ 878; для совокупности N_A и $N_В$—905; для совокупности $N_Б$ и $N_В$—844. Доля m_1 равна, соответственно: 6,7%, 6,9%, 6,9%.

Напомним, что доля m_1 в совокупной выборке составляет 5,5%. Нас интересует также, какова эффективность словарей, составленных по выборкам А, Б, В (т. е. степень покрываемости каждым из словарей (L_A, $L_Б$, $L_В$) новой выборки — в данном случае из числа имеющихся).

Теоретически эффективность исследуемого словаря оценивается по формуле, позволяющей определить долю заполненного данным словарем текста (иначе—величину коэффициента заполнения ζ):

$$\zeta = 1 - \frac{m_1}{N}, \qquad (5)^{22}$$

где N — объем выборки, по которой составлен исследуемый словарь, m_1 — количество однократных словоупотреблений в данной выборке. Теоретические значения коэффициента заполнения[23] (ζ) для исследуемых словарей L_A, $L_Б$, $L_В$ следующие:

Таблица 8

L	N	m_1	ζ
L_A	7005	635	0,91
$L_Б$	5985	566	0,906
$L_В$	6161	604	0,902

Экспериментальные значения к. з. (ζ), показывающие, как исследуемый словарь реально заполняет новую выборку[24], таковы:

---

[22] Формулу (5), $\zeta = 1 - \frac{m_1}{N}$ выводят из формулы $q = m_1 \frac{N_0}{N}$, с помощью которой для некоторой выборки (или словаря) N определяют прирост новых словоформ или лексем (q) из выборки (или словаря) N_0, в предположении, что эти новые словоформы (или лексемы) при наличии достаточно большого объема имеющейся выборки (или словаря) N будут редкими и, следовательно, однократными (Е. А. Калинина, Изучение лексико-статистических закономерностей на основе статистической модели («Статистика речи», Л., 1968, стр. 75—76).

[23] Далее—сокращенно: к. з.

[24] L_A — выборку Б, выборку В; $L_Б$ — выборку А, выборку В; $L_В$ — выборку А, выборку Б.

Таблица 9

Выборки	ζ		
	L_A	$L_Б$	$L_В$
А		0,883	0,876
Б	0,892		0,888
В	0,887	0,897	

Область пересечения попарно взятых словарей, обусловливающая то или иное покрытие соответствующей выборки, охватывает: при наложении L_A на $L_Б$ — 572 лексемы; при наложении L_A на $L_Б$ — 556 лексем; при наложении $L_Б$ на $L_В$ — 563 лексемы. Таким образом, примерно 1/2 часть каждого из исследуемых словарей покрывает не менее 87% текста новой выборки.

Это означает, что не входящие в область пересечения исследуемых словарей лексемы — низкочастотные, хотя и (по данным, полученным эмпирически) не все с лишь однократным употреблением[25]. Близость значений к. з., полученных для исследуемых словарей, вновь свидетельствует об упорядоченности и однородности статистической структуры телеграфных текстов и, кроме того, позволяет еще сформулировать следующее предположение: словарь лексем, составленный по выборке N, имеет высокую и примерно одинаковую степень покрываемости разных выборок N_1, N_2, N_3...Nп (равного с N объема)[26]. Сказанное приводит к выводу, что степень покрываемости данным словарем всей совокупности указанных выборок ($N_1+N_2+...N_п$) и степень покрываемости им любой из указанных выборок также примерно одинаковы (точнее: степень покрываемости данным словарем всей совокупности указанных выборок есть среднее арифметическое значений к. з. каждой из выборок, составляющих совокупность).

Полученные для каждого из исследуемых словарей (L_A, $L_Б$, $L_В$) экс-

---

25 Предположение «новая для исследуемого словаря лексема—лексема с однократным употреблением» оправдывается не в полной мере, поскольку исследуемые словари равны и недостаточно велики по объему (см. сноску 1); тем не менее, распределение по частотам лексем, входящих и не входящих в область пересечения сопоставляемых словарей, соответствует указанному предположению по тенденции.

26 Удовлетворительное совпадение теоретического и экспериментальных значений к. з., полученных для каждого из исследуемых словарей, как будто подтверждает правильность сделанного предположения; однако его следует проверить более тщательно, особенно если выборки N и N_1, к примеру, будут резко отличны по фактору времени или по тематическому признаку; также может возникнуть необходимость установить наименьший объем выборки, при котором сделанное предположение останется справедливым.

периментальные значения к. з. новой совокупности в 2000 телеграмм[27] служат подтверждением данного вывода:

Таблица 10

Совокупность выборок	ζ		
	L_A	$L_Б$	$L_В$
А + Б			0,882
А + В		0,890	
Б + В	0,890		

У нас имеется возможность определить эффективность словаря, составленного по выборке в 2000 телеграмм.

Теоретические значения к. з. для указанных словарей ($L_{АБ}$, $L_{АВ}$, $L_{БВ}$) следующие:

Таблица 11

L	N	m_1	ζ
$L_{АБ}$	12990	861	0,938
$L_{АВ}$	13106	890	0,933
$L_{БВ}$	12146	841	0,931

Экспериментальные значения к. з. для исследуемых словарей (при покрытии ими новой выборки в 1000 телеграмм) равны:

Таблица 12

Выборки	ζ		
	$L_{АБ}$	$L_{АВ}$	$L_{БВ}$
А			0,910
Б		0,921	
В	0,922		

[27] Здесь, как и всюду в данной работе, все эксперименты производятся в пределах имеющегося материала в 3000 телеграмм, почему и отсутствует возможность далее увеличивать размер совокупности, на которой испытывается исследуемый словарь.

Область пересечения каждого из исследуемых словарей ($L_{АБ}$, $L_{АВ}$, $L_{БВ}$) со словарем новой выборки в 1000 телеграмм охватывает: при наложении $L_{АБ}$ на $L_{В}$ — 677 лексем; при наложении $L_{АВ}$ на $L_{Б}$ — 693 лексемы; при наложении $L_{БВ}$ на $L_{А}$ — 686 лексем. Следовательно, примерно 40% лексем в каждом из исследуемых словарей[28] покрывает не менее 91% новой для исследуемого словаря выборки в 1000 телеграмм.

При увеличении размеров указанных словарей до размеров словаря сводной выборки в 3000 телеграмм ($L_{АБВ}$ — 2031 лексема) область пересечения словарей выборок А, Б, В охватывает уже 807 лексем (соответствующее число словоупотреблений — 16309 из 19151).

Теоретическое значение к. з. для словаря сводной выборки (которое только и может быть вычислено) оказывается равным 0,945 (m_1 — 1054, N — 19151). Как же велик должен быть по объему словарь лексем, заполняющий любой новый текст на 99%?

Определим объем выборки, позволяющей составить такой словарь (по-видимому, полный и достоверный). Если значение к. з. для указанного словаря принято равным 0,99, то доля однократных словоупотреблений (m_1) должна составить 0,01 (в соответствии с формулой (5): $\zeta = 1 - \frac{m_1}{N}$).

Подставив эти значения в формулу для определения искомого объема выборки $N = \frac{4q}{\delta^2 p}$ (где p—доля m_1, q—доля остальных словоупотреблений, δ — относительная ошибка, равная в данном случае 0,05)[29], найдем: искомый объем выборки — 158000 словоупотреблений (что соответствует примерно 20000 телеграмм)[30].

Целесообразно, исследуя этот массив, разбить его на тематические группы. Тогда кривая частоты каждой лексемы при сведении воедино тематических (но построенных с учетом фактора времени) словарей позволит заметить нюансы в статистической характеристике лексемы (стабильность; периодическое появление — исчезновение и т. п.).

О том, насколько специфичен отбор лексем в телеграфных текстах, можно судить, рассматривая группу слов с i (рангами) 1—41 в частотном микрословаре телеграфных текстов, в словаре Э. А. Штейнфельдт[31] и в частотном словаре (словоформ!) подъязыка электроники[32].

---

[28] $L_{АБ}$ — 1634 лексемы,
$L_{АВ}$ — 1655 лексем,
$L_{БВ}$ — 1586 лексем.

[29] Б. Н. Головин, указ. соч., стр. 59.

[30] Которые следует собирать, разумеется, в течение года—от сезона к сезону.

[31] Э. А. Штейнфельдт. Частотный словарь современного русского языка, Таллин, 1963.

[32] Е. А. Калинина, Частотный словарь русского подъязыка электроники (в сб. «Статистика речи», Л., 1968, стр. 144—150).

Таблица 13[33]

	Частотный микрословарь подъязыка телеграмм	Словарь Э. А. Штейнфельдт	Словарь подъязыка электроники
Со:	И	И, что, а, как, но	И, что, как, а, или, если, то(?), чем(?)
Пр	В (ВО), (НА), С (СО)	В (ВО), НА, С (СО), к (ко), у, за, по, из (изо), о (об), от	В, при, НА, для, с, от, по, из, к, до, между, через
	(НЕ), срочно, крепко, благополучно	НЕ, же (ж), так, вот, только, еще, то	НЕ так, где, же,
Н[34]	*весь*, дорогой	этот, *весь*, тот, свой, один, который, наш, такой	это (?)
С	день, вагон, здоровье, поезд, рождение, счастье, *ты-вы*, письмо, июнь, деньги, успех, каждый-все, час, жизнь, *я-мы*, востребование, привет, главпочтамт	он, она, они, *я, мы, ты, вы*.	рис., тока, ток, напряжения, электронов, случае, поля, времени, атом, напряжение, катода, время, области, разряда
Гл.	Целовать, поздравлять, желать, встретить — встречать, выехать—выезжать, выслать—высылать, *быть (есть)*, приехать — приезжать, ждать, просить, сообщить — сообщать, получить—получать, телеграфировать.	*Быть (есть)*, сказать, мочь, говорить	Можно, может

Отметим, что шесть слов из рассмотренной группы в частотном микрословаре телеграмм (выслать-высылать, благополучно, востребование, привет, главпочтамт, телеграфировать) вообще отсутствуют в словаре Э. А. Штейнфельдт.

Ознакомление с частотным микрословарем телеграмм (и сопоставление его — по рангам $1 > 41$ — со словарем Э. А. Штейнфельдт)[35] позволит убедиться в том, что и в остальной части частотный микрослo-

---

[33] Прописными буквами набраны слова, общие для всех трех словарей, курсивом — слова, общие для словаря Штейнфельдт и микрословаря подъязыка телеграмм; черным шрифтом—слова, общие для словаря Штейнфельдт и словаря электроники.

[34] В словаре Штейнфельдт классификация частей речи — традиционная: частицы (как и местоимения) образуют самостоятельный класс слов.

[35] Ранговую корреляцию по соответствующей формуле определять в данном случае неуместно, ввиду малых размеров выборки и небезусловности данных в частотном микрословаре телеграмм (напомним, что и в более благоприятных случаях формулу для определения коэффициента ранговой корреляции следует применять с осторожностью).

варь телеграмм демонстрирует значительное отличие (или тенденцию к отличию!) в отборе и употребительности лексем в телеграфных текстах (по сравнению с отбором и употребительностью лексем в литературном русском языке).

ՀԵՌԱԳՐԱՅԻՆ ՏԵՔՍՏԵՐԻ ՎԻՃԱԿԱԳՐԱԿԱՆ ԿԱՌՈՒՑՎԱԾՔԻ ՆԿԱՐԱԳՐՈՒԹՅՈՒՆԸ

ՆԱՋԻԿ ՄԵԼԻՔՅԱՆ

Ա մ փ ո փ ո ւ մ

Հեռագրային տեքստերի վիճակագրական կառուցվածքի քննության համար նյութ են ծառայել քերականության և բառապաշարի քանակական տվյալները, որոնք սացվել են մեր կողմից կազմած հեռագրերի հաճախական միկրոբառարանից։

Ընարանքի առանձին, մոտավորապես հավասար շերտերից (երեք հազար հեռագրերից յուրաքանչյուր հազարի սահմաններում) սացված տվյալների համագրումը հնարավորություն է տալիս եզրակացնելու հեռագրերի վիճակագրական կառուցվածքի համասեռության մասին։ Հավասար ծավալ ունեցող ընտրանքների հիման վրա գազմված բառարանների էֆեկտիվությունը արտահայտում է մոտավորապես նույն արժեքներով (ինչպես տեսական, այնպես էլ փորձառական)։

Հեռագրային տեքստերի յուրաքանչյուր նոր ամրողջության 99%-ը ծածկող հավաստի հաճախական բառարան կազմելու համար անհրաժեշտ է ընտրանք, որն ընդգրկում է մոտավորապես քսան հազար հեռագիր։

Ուսումնասիրության արդյունքները ապացուցում են հեռագրային տեքստերի հատկարգային բնույթը և արգարացնում նրանց հետազոտումը լեզվավիճակագրության մեթոդներով։