

УДК 414:0.093

## О НЕКОТОРЫХ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИКАХ ПРЕФИКСОВ

Н.В.Саломатина, Л.С.Юдина

В обзоре лингвистических работ, утвержденных коллегией ВАК СССР в 1981-1985 гг., отмечается "острая необходимость исследований в области общения человека с ЭВМ" [1]. Речь идет об изучении внутренней организации языковой системы; при этом на первом плане - выделение основных единиц языка на каждом из его уровней, освещение различных аспектов взаимозависимости и взаимодействия уровней.

В этой связи особая роль отводится в настоящее время исследованию морфемного представления слова и собственно морфемы - минимальной значимой единицы языка [2]. Слово, с одной стороны, обладает фонетической целостностью, а с другой - грамматической структурированностью. Морфемная структура содержит все системные связи данного слова, фиксирует все его грамматические возможности; морфемная структура - универсальный показатель согласования. Кроме того, предельная членимость и регулярная воспроизводимость морфем, их довольно твердая привязка к определенному месту в слове позволяет в задачах, связанных с автоматической обработкой текстов, оперировать (с помощью ЭВМ) практически неограниченным словарем, представленным в имплицитной форме. Присвоив морфеме соответствующие значения на основе ее фонетических характеристик, ее можно использовать, в частности, для адекватной фонетической интерпретации слабых участков речевой цепи в задачах распознавания.

Однако, чтобы использовать морфему в качестве оперативной единицы словаря распознавания, необходимо располагать статистически достоверными данными о самом корпусе морфем, об их соотношениях и внутрисловных окружениях для любой морфемы.

По мнению специалистов, "применительно к русскому языку известны некоторые работы в этом направлении, но, к сожалению, они

мание исследователей привлекли только корневые единицы. Аффиксальные же морфемы рассматривались очень непоследовательно, а иногда и неправильно. Вот почему ни одна из существующих работ не содержит исчерпывающего списка морфологических единиц русского языка" [3].

Положение дел в этой области таково, что хотя к настоящему времени и разработана довольно солидная "машинная морфология" русского языка, "практически сегодня каждая конкретная автоматизированная информационная система использует свой алгоритм морфологического анализа, а многие информационные системы работают вообще без привлечения морфологии, минуя морфологию" [4], на базе проблемно-ориентированных словарей.

При переходе к большим (неограниченным) словарям такой подход становится неприемлемым, и появляется необходимость в обращении к морфологии. Число морфем, по крайней мере, на порядок меньше, чем число слов, и на два — чем число словоформ. Представляется, что привлечение морфологии позволит наиболее экономным образом использовать морфему при содержательном анализе текста — для верификации внутрисловных структур (введенный в ЭВМ словарь позволяет оперировать всеми возможными внутрисловными контекстами для каждой морфемы).

В предлагаемой статье речь идет об одном из видов аффиксальных морфем — о префиксах (приставках), рассматриваются их количественные и качественные характеристики. Работа является продолжением исследований с помощью ЭВМ морфемной структуры русского языка на материале словообразовательного словаря\*) объемом свыше 100 тыс. слов [5].

Полученные на материале словаря статистические закономерности, касающиеся собственно словообразования, а именно: позиционное распределение приставок по отношению к корню и их связь с ча-

\*) Слова в словаре представлены в виде цепочек морфов, разделенных пробелами. Членение слов на морфемы производилось автоматически. При этом все внутрисловные элементы были разделены на три категории: докорневые, корневые и послекорневые. В результате докорневые элементы включали в свой состав префиксы, префиксоиды, вторые корни с их аффиксами. Авторами данной работы при записи словаря на ленту были внесены соответствующие коррективы. Однако сугубо этимологический подход авторов словаря к членению слов обусловил появление отдельных, не характерных для современного русского языка морфов, которые, впрочем, не могут помешать выявлению общих морфологических тенденций.

стеречным значением, являются общезыковыми и при переходе к текстам существенно не изменяются.

Данные по статистике префиксов получены для трех частей речи - существительных, прилагательных и глаголов, как наиболее конструктивных членов любого предложения. Из рассмотрения исключены сложные слова (~14%), а также прочие части речи (тоже ~14%). Таким образом, материалом для анализа послужило более 72 тыс. слов словаря, в том числе существительных ~30 тыс., прилагательных ~16 тыс., глаголов ~26,5 тыс.

Т а б л и ц а I

Части речи	Без приставок, %	С приставками, %
Существительные	73	27
Прилагательные	69	31
Глаголы	18	82

Представляет интерес количественное соотношение бесприставочных слов и слов, имеющих приставки, как между частями речи, так и внутри каждой из них (см. табл. I). Как видно из

табл. I, среди рассмотренных частей речи более всего префиксы характерны для глаголов и менее всего - для существительных.

Следует сказать несколько слов о функциональных особенностях приставок, которые, будучи приняты во внимание, могут оказаться полезными при распознавании речи. Как известно, основное назначение приставок - словообразование. Присоединение приставки не изменяет стеречное значение слова; для приставки не обязательна тесная связь со словами какой-либо одной части речи, например: о-зарить, в-зарение, о-заренный. Следует особо отметить обособленность, структурную самостоятельность префиксов по сравнению с суффиксами - приставки не зависят от флексий и не несут информации о грамматических связях слова. Префиксы реже влияют на строение начала слова, чем суффиксы на строение конца основы.

О бóльшей отъединенности приставок от корня в слове свидетельствуют также следующие факты:

- приставки могут иметь побочные ударения;
- позиционные изменения звуков, обязательные для всех морфем в слове, могут не затрагивать приставок;
- ассимилятивное смягчение согласных на стыке приставки и корня тоже происходит иначе, чем на стыке корня и суффикса;
- на стыке приставки и корня отсутствуют процессы наложения (например: со-ответчик).

Эта явно выраженная универсальность значения приставок и их структурная самостоятельность определяют высокую продуктивность (словообразовательную способность) приставок.

Известно, что приставки встречаются как одиночные, так и в сочетаниях. Оди н о ч н ы е приставки составляют ~90%. По частям речи они распределяются неравномерно: у существительных - 22%, у прилагательных - 15%, у глаголов - 63%.

Всего одиночных приставок в словаре 32,3 тыс., р а з н ы х среди них ~220 штук. Обращает на себя внимание неравномерность распределения частот. Первые 25 префиксов частотной таблицы, имея частоту встречаемости среднюю (0,5%) и выше, составляют ~90% от всех встретившихся. На рис. I приведен график частоты встречаемости наиболее продуктивных приставок.

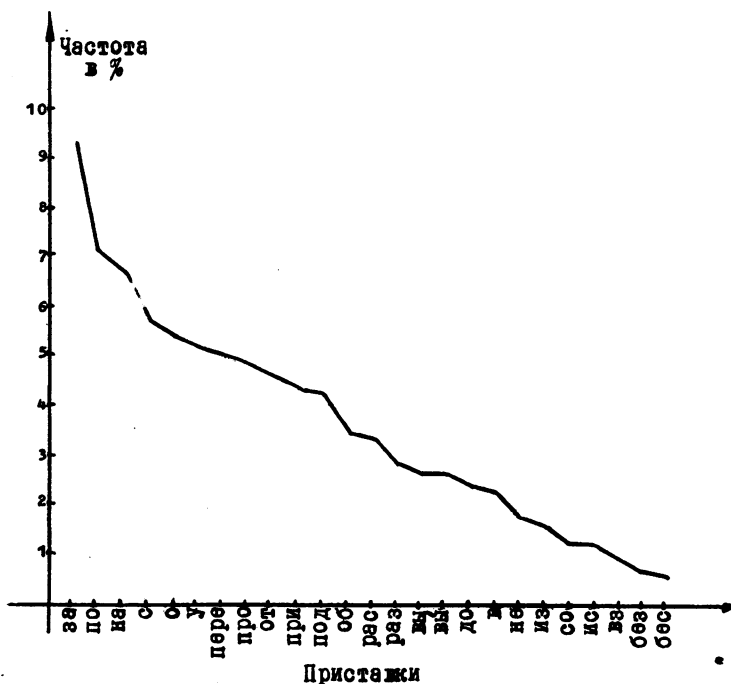


Рис. I

При этом 13 самых частых из них (см.рис.1) имеют очень высокие – четырехзначные значения частоты встречаемости и участвуют в образовании почти 70% слов исследуемой части словаря.

Однако к концу третьего десятка частотной таблицы эти значения стремительно понижаются и вскоре вырождаются в длинный шлейф с низкими (в интервале от 0,5 до 0,003%) частотами. Таких низкочастотных приставок около 200 штук.

Наличие столь большого числа низкочастотных префиксов объясняется, с одной стороны, самим характером исследуемого материала – словарь отличается от любого текста большим разнообразием лексики; с другой стороны, это результат излишне дробного членения на морфемы.

Однако эти особенности не затемняют основных выводов, и полученные результаты хорошо согласуются, например, с описанными в

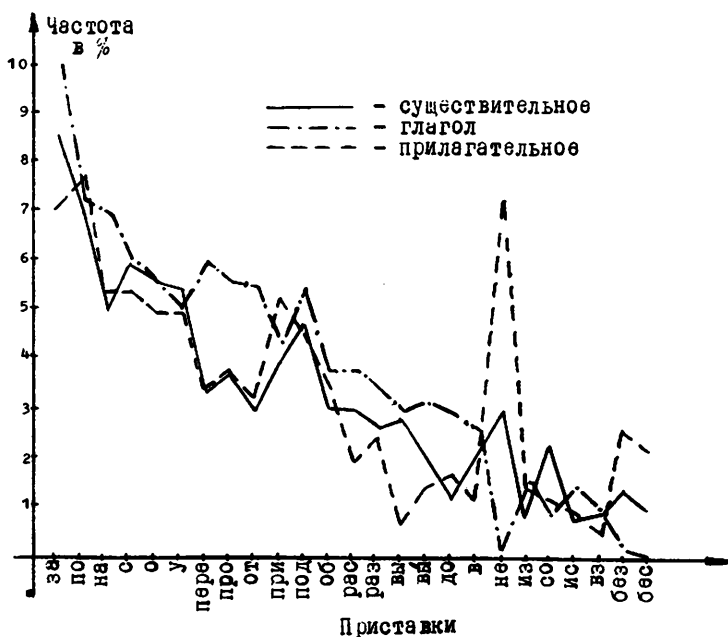


Рис. 2

литературе [6,7], одни из которых (см. [6]), кстати, получены по текстам. Напомним, что все сказанное относится к суммарной частоте встречаемости префиксов.

Нами получены также значения частоты встречаемости префиксов по частям речи. На рис.2 приведен график распределения частот встречаемости наиболее продуктивных префиксов в зависимости от частеречного значения. При общей схожести кривых они имеют различия, говорящие о том, что, например, приставки (за-, на-, пере-, про-, от-, рас-, раз-, вы-) более характерны для глаголов, а приставки (не-, по-, под-, со-, без-, бес-) - для имен существительных и прилагательных.

В табл.2 приведены абсолютные частоты встречаемости наиболее продуктивных префиксов отдельно для существительных, прилагательных и глаголов, что позволяет получить представление о роли конкретной приставки в функционировании данной части речи.

Представляют также интерес характеристики одиночных префиксов по СГ-структурам и содержание этих структур. (Напомним, что одиночные приставки составляют ~90%.)

Среди одиночных приставок встретилось:

- однобуквенных 19%, в том числе структур С - 8%, Г - 11%;
- двухбуквенных 50%, из них структур СГ - 36%, СС - 2%, ГС - 12%;
- трехбуквенных 25%, где СГС - 14%, ССГ - 10%, ГСС - 0,5%, ГСГ - 0,5%;
- четырехбуквенных 6%, в том числе СГСГ - 5%, ССГС - 1%.

Кроме того, нам представлялось интересным получить СГ-характеристики начал корней, следующих непосредственно за приставкой, с тем, чтобы проверить, как выдерживается основной принцип слогаделения - по открытым слогам [8,9] на границе морфем.

Ниже приведены схемы, отражающие сочетаемостные тенденции СГ-структур на границе префиксов и корней<sup>х)</sup>.

х)

На схемах: слева - приставка, справа - начало корня, далее - частота встречаемости данной последовательности, абсолютная и в процентах.

Т а б л и ц а 2

Частота встречаемости префиксов по частям речи

№ п/п	Существительное		Прилагательное		Глагол	
	Префикс	Абсолют. частота	Префикс	Абсолют. частота	Префикс	Абсолют. частота
1	2	3	4	5	6	7
1	за-	612	по-	363	за-	2062
2	по-	494	не-	344	по-	1463
3	с-	423	за-	336	на-	1402
4	о-	399	на-	256	пере-	1214
5	у-	386	с-	254	с-	1190
6	на-	369	при-	252	про-	1144
7	под-	344	у-	244	о-	1144
8	при-	287	о-	243	от-	1129
9	про-	268	под-	216	у-	1057
10	пере-	240	про-	182	при-	881
11	не-	220	об-	164	под-	839
12	от-	217	пере-	160	рас-	775
13	об-	216	от-	158	об-	770
14	рас-	212	без-	122	вы-	661
15	вы-	210	раз-	120	раз-	637
16	раз-	197	бес-	107	вы-	629
17	со-	163	рас-	102	до-	599
18	в-	145	до-	87	в-	536
19	вы-	137	из-	72	из-	353
20	без-	98	вы-	71	ис-	315
21	из-	95	со-	59	вз-	230
22	до-	86	в-	57	со-	213
23	бес-	78	ис-	48	вс-	134
24	вз-	67	вы-	46	воз-	127
25	де-	59	над-	46	вос-	77
26	ис-	59	пред-	37	над-	76
27	ре-	54	пре-	35	пре-	62
28	пред-	53	а-	31	обо-	49
29	пре-	48	воз-	24	во-	46
30	а-	47	вз-	23	де-	45
31	воз-	45	ре-	23	пред-	35
32	по-	38	су-	22	подо-	30
33	вс-	36	вне-	20	разо-	30

Продолжение таблицы 2

1	2	3	4	5	6	7
34	над-	33	вос-	15	ото-	30
35	за-	32	при-	14	разъ-	26
36	вос-	31	ин-	14	объ-	25
37	о-	31	по-	14	ре-	24
38	су-	30	за-	12	изъ-	19
39	при-	26	де-	12	низ-	13
40	про-	25	про-	10	съ-	13
41	ин-	24	подъ-	9	дез-	12
42	па-	23	на-	9	бес-	12
43	ан-	23	па-	9	не-	11
44	ди-	22	су-	9	възъ-	11
45	от-	17	вс-	9	изо-	11
46	не-	116	объ-	8	без-	9
47	су-	15	ан-	8	су-	9
48	под-	14	со-	8	отъ-	9
49	обо-	14	би-	8	вн-	8
50	об-	14	обо-	7	сн-	8
51	объ-	13	ди-	7	нис-	7
52	би-	13	разъ-	7	ин-	7
53	во-	13	вн-	6	подъ-	6
54	ак-	12	о-	6	надо-	6
55	им-	12	подо-	5	взо-	6
56	на-	9	подо-	5	ко-	5
57	дез-	10	ото-	5	ра-	4
58	пра-	9	дис-	5	въ-	4
59	рос-	9	изъ-	5	а-	4
60	пра-	9	пра-	4	въ-	4
61	дис-	8	об-	4	экс-	4
62	роз-	7	до-	4	предъ-	3
63	пара-	7	обо-	4	дис-	3
64	подъ-	6	во-	4	па-	2
65	син-	6	въ-	4	со-	2
66	ко-	6	низ-	4	за-	2
67	из-	6	ир-	4	под-	2
68	вне-	5	безъ-	4		
69	ир-	5	съ-	4		
70	воз-	3	нис-	4		

1) Г <— СГ - 2433 (67%)  
           — СС - 1148 (32%)  
           — ГС - 39 (1%)

С <— СГ - 2023 (76%)  
           — СС - 539 (21%)  
           — ГС - 83 (3%)

СГ <— СГ - 7662 (66%)  
           — СС - 3390 (30%)  
           — ГС - 465 (4%)

2) СС <— СГ - 362 (66%)  
           — СС - 114 (21%)  
           — ГС - 73 (13%)

ГС <— СГ - 2965 (76%)  
           — СС - 653 (16%)  
           — ГС - 304 (8%)

3)

СГ <— СГ - 3256 (70%)  
           — СС - 1087 (23%)  
           — ГС - 284 (7%)

ССГ <— СГ - 2244 (68%)  
           — СС - 919 (28%)  
           — ГС - 120 (4%)

ГСС --- ГС - 20  
           ГСГ --- СС - 113

СГСГ <— СГ - 1113 (65%)  
           — СС - 524 (30%)  
           — ГС - 65 (5%)

4)

ССГС <— СГ - 90 (67%)  
           — СС - 33 (25%)  
           — ГС - 11 (8%)

Судя по приведенным данным, независимо от качества приставки следующий за ней корень в 95-98% случаев начинается с согласной. Это обстоятельство, обнаруживающееся при наложении морфемной сегментации на фонемную, прежде всего еще раз говорит о структурной самостоятельности, об автономности префиксов.

Относительно конкретных гласных и согласных элементов, входящих в состав приставок, следует сказать, что число их ограничено и сочетаются они друг с другом избирательно. Из согласных наиболее часты в составе приставок щелевые звонкие и глухие [з, в, с], смычные глухие и звонкие [п, б, д, т], сонанты [н, р].

Число гласных, в сочетании с которыми эти согласные образуют приставку<sup>ж)</sup>, тоже невелико; это, в основном, открытые (широкие) нижнего и среднего подъема [а, о, е]; далее - закрытый верхнего подъема [ы]. Редки приставки с [и, у].

Перечисленные согласные и гласные образуют как одно-, так и многобуквенные приставки; в первом случае это суть согласные и гласные, во втором - строго определенные комбинации из них же.

Таким образом, встретившиеся в словаре в качестве приставок буквосочетания (эн-, аф-, ал-, ил-, чу-, ом-, ага-, ...), а также (ере-, ду-, ба-, ун-, аб-, тир-, ...) по формальным признакам не

ж)

С согласной начинается 76% приставок.

могут быть причислены к приставкам: первые включают "чужие" буквы, вторые - из-за нарушения в комбинациях (таких сочетаний среди русских приставок нет, хотя каждая в отдельности эти графемы входят в их состав).

Иноязычных приставок или начальных частей приставочного типа (а-, анти-, дез-, дис-, син- и др.) среди одиночных насчитывается ~2% - 700 штук всего, 48 разных; в цепочках их единицы.

Все сказанное относится к одиночным префиксам.

Встречаются также префиксы в сочетаниях (цепочках). Всего таких цепочек ~3,3 тыс. - это 10% от всех слов, образованных при помощи приставок. Разных среди них свыше 500 штук (для сравнения: одиночных разных ~220 штук). По частям речи цепочки распределены так: существительных с такими префиксами - 26%, прилагательных - 27%, глаголов - 47%.

Последовательности из двух приставок составляют ~4%, из трех ~6%. Встретившиеся (единичные) максимально длинные префиксальные последовательности содержат 4 элемента (например, не-при-с-по-собленность), занимающие соответственно 4 позиции перед корнем; отсчет ведется от корня.

Табл.3 дает представление о наиболее частых сочетаниях приставок. Всего разных сочетаний из двух приставок свыше 450 штук, из трех ~50 штук. Приведенные 35 самых частых двух- и трехэлементных приставок составляют от всех двухэлементных 40%, а от трехэлементных 70%.

В последовательностях префиксов для нас представляют интерес фонетические характеристики приставок и частота встречаемости их по позициям по отношению к корню.

Табл.4 отражает обобщенные характеристики приставок по числу букв и СГ-структурам в связи с позицией перед корнем.

Анализ приставок, входящих в цепочки, по числу букв показывает, что наиболее распространенные (двух- и трехбуквенные структуры, в сумме они составляют 70%) встречаются во всех трех позициях почти с одинаковой частотой. В этом отношении более показательны редкие - одно- и четырехбуквенные структуры: однобуквенные располагаются ближе к корню, четырехбуквенные - дальше от корня.

При сравнении префиксальных СГ-структур в цепочках как внутренних позиций, так и по позициям, видно, что по мере удаления от корня растет число структур СГ и ССГ, что, очевидно, является отражением принципа организации речевого потока по открытым слогам.

Т а б л и ц а    3

Наиболее частые сочетания префиксов

№ п/п	Двухэлементные		Трехэлементные	
	Префикс	Абсол. частота	Префикс	Абсол. частота
I	не-до-	132	при-с-по	12
2	не-у-	90	о-без-за-	9
3	о-без-	66	не-до-вы	7
4	не-по-	63	пред-о-пре-	7
5	о-бес-	59	пред-рас-по-	6
6	не-о-	53	пере-о-с	6
7	не-с-	43	вос-про-из	6
8	не-при-	42	о-бес-с-	5
9	при-у-	39	не-про-из-	5
10	о-с-	36	не-у-по-	5
11	не-за-	35	пере-рас-пре-	5
12	пере-с-	31	не-до-у	4
13	раз-у-	31	не-пред-у	4
14	о-по-	30	при-со-во-	4
15	рас-по-	30	не-без-у-	4
16	пере-у-	28	не-до-с-	4
17	из-об-	26	не-пре-о-	3
18	пред-у-	26	пере-про-из-	3
19	по-с-	26	не-со-в-	3
20	у-с-	26	о-без-на-	3
21	про-из-	25	не-рас-по-	3
22	с-по-	25	не-до-вы-	3
23	не-ис-	25	не-пред-на-	2
24	пере-на	24	по-рас-с	2
25	пере-о-	24	о-без-об-	2
26	без-у-	24	не-до-раз-	2
27	не-об-	22	на-без-об-	2
28	без-от-	22	не-со-раз-	2
29	при-о-	22	не-со-из-	2
30	пред-по-	22	не-без-рас-	2
31	не-про-	22	не-бес-с	2
32	не-раз-	22	не-о-про-	2
33	раз-о-	21	не-бес-по-	2
34	рас-с-	21	не-о-пре-	2
35	не-от-	21	не-из-рас-	2

Т а б л и ц а 4

Распределение различных типов префиксов  
по позициям перед корнем

Префиксы	Структура префикса	I позиция	II позиция	III позиция
Однобуквенные	С	14	11,5	-
	Г	17	2,5	15
		31%	14%	15%
Двухбуквенные	СГ	36	42	54
	СС	0,03	0,4	-
	ГС	12,47	4,6	1
		48,5%	47%	55%
Трехбуквенные	ССГ	8,5	9,5	14
	СГС	11,5	17,47	4
	СГГ	-	0,03	-
		20%	27%	18%
Четырехбуквенные	СГСГ	0,4	7,6	4,5
	ССГС	0,1	4,3	7,5
	ГССГ	-	0,1	-
		0,5%	12%	12%

Возможно, на этом основании для практических целей, связанных с автоматической обработкой текстов, последовательность приставок в структурном отношении допустимо рассматривать как единое целое, находящееся слева от корня.

Тем не менее рассмотрим подробнее фонетические характеристики приставок в зависимости от распределения их по позициям.

В I позиции встретилось ~50%, во II - ~40%, в III - 13 р а з н ы х префиксов. Различные (по качеству и по количеству) комбинации этих префиксов составляют ~3,3 тыс. цепочек, образующих ~90% слов с этим видом приставок. В целях исключения случайных ошибок, возникших при записи словаря на магнитную ленту, а также в результате автоматического членения слов на морфемы одиночные цепочки типа: су-по-, ваз-о- и пр. (они составляют ~10%) не рассматривались.

На рис.3 приведено распределение частоты встречаемости префиксов, входящих в цепочки, в зависимости от места перед корнем.

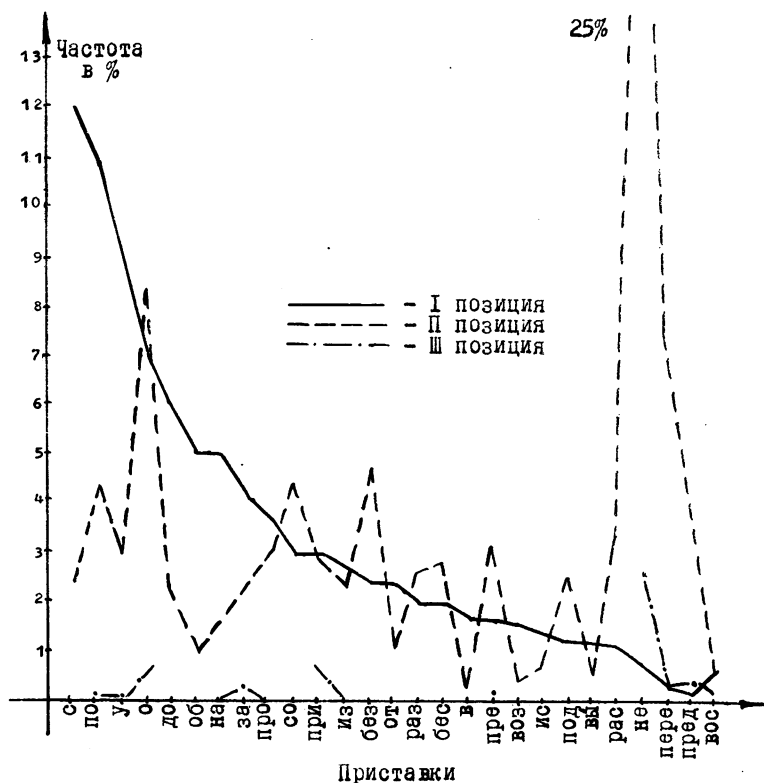


Рис. 3

Табл.5 позволяет сравнить абсолютную частоту встречаемости префиксов как одиночных, так и в сочетаниях. Приведенные 30 префиксов образуют свыше 95% всех слов с приставками, соответственно с одной или более.

Легко заметить, что наиболее частые одиночные префиксы являются наиболее частыми и в сочетаниях. Однако очевидно также, что существует, во-первых, зависимость от того, как встречается приставка — одна или в сочетании, и, во-вторых, некоторая закреплённость префиксов за позицией.

Т а б л и ц а   5

Сравнительная таблица абсолютной частоты  
встречаемости префиксов

№ п/п	Одиночные		В с о ч е т а н и я х					
	Префикс	Число слов с ним	I позиция		II позиция		III позиция	
			Префикс	Число слов с ним	Префикс	Число слов с ним	Префикс	Число слов с ним
I	за-	3010	с-	379	не-	745	не-	79
2	по-	2320	по-	334	о-	248	о-	22
3	на-	2027	у-	277	пере-	224	при-	16
4	с-	1867	о-	214	без-	140	пред-	13
5	о-	1786	до-	175	по-	132	за-	9
6	у-	1687	об-	157	со-	132	пере-	8
7	пере-	1614	на-	155	пред-	114	вос-	6
8	про-	1594	за-	126	рас-	102	пре-	6
9	от-	1504	про-	110	пре-	97	по-	4
10	при-	1420	со-	84	у-	90	у-	4
11	под-	1399	при-	83	про-	90	про-	3
12	об-	1150	из-	80	при-	89	на-	2
13	рас-	1089	без-	72	бес-	84	из-	1
14	раз-	954	от-	72	раз-	79		
15	вы-	885	раз-	64	под-	76		
16	вы-	869	бес-	62	с-	75		
17	до-	772	в-	51	за-	73		
18	в-	738	пре-	50	из-	72		
19	не-	575	воз-	48	до-	66		
20	из-	520	ис-	42	на-	48		
21	со-	435	под-	36	об-	32		
22	ис-	422	вы-	35	от-	30		
23	вз-	320	рас-	34	вос-	21		
24	без-	229	су-	23	ис-	20		
25	бес-	197	по-	23	вы-	16		
26	воз-	196	не-	23	вы-	15		
27	вс-	179	во-	21	воз-	13		
28	над-	155	вы-	20	вс-	12		
29	пре-	145	вос-	19	вс-	11		
30	пред-	125	ко-	19	нис-	10		

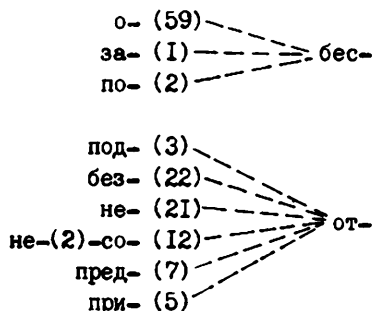
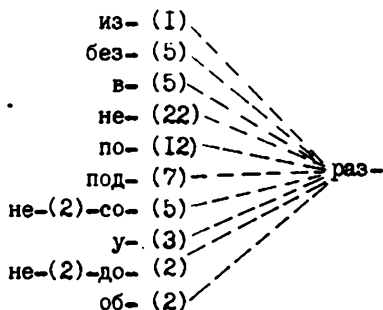
Кроме количественных характеристик префиксов в сочетаниях, представляет интерес также число ветвлений, отражающее способность той или иной приставки к образованию разных цепочек. По-видимому, прямой зависимости между частотой встречаемости приставки и числом ветвлений нет. Сравним, например, следующие приставки:

	Абсолютная частота	Число ветвлений
с-	379	19
по-	334	23
у-	277	11
без-	72	2
бес-	62	3
по-	23	13

Малое число ветвлений при высокой частоте встречаемости приставки означает высокую нагруженность одной (или нескольких) ветвей. Следует заметить, что нагруженность между ветвями распределена очень неравномерно. Ниже приведены примеры ветвлений, реализующие соединение приставок в цепочки. Для иллюстрации взяты приставки:

раз-	частота 64, число ветвлений 10,
от-	частота 72, число ветвлений 6,
бес-	частота 62, число ветвлений 3.

На схеме: префикс справа – общая часть всех цепочек, стоящая рядом с корнем. Префиксы и цифры слева конкретизируют начала цепочек по качеству и количеству.



Очевидно, сочетаемость префиксов, как и линейная организация цепи морфов в слове, определяется не только сегментными, но и суперсегментными условиями. В общем случае и структура русских морфем, и модели русского словообразования обеспечивают соответствие формы производного слова требованиям всех уровней языка.

Линейные взаимодействия в цепи морфов обусловлены и фонетически, и морфологически. В дальнейшем в целях исследования соотношения фонемной сегментации слова с морфемной представляет интерес провести сравнение полученных ранее [10] характеристик консонантных начал слов с предлагаемыми статистическими характеристиками префиксов.

Практическое значение выделения как типовых, так и конкретных морфологических структур заключается в более глубоком проникновении в построение слова. Это дает возможность при решении прикладных задач (например, автоматическое индексирование, реферирование, автоматическое исправление ошибок, т.е. везде, где используется морфемный анализ) опираться не только на теорию, но и на реальный материал.

Таким образом, в результате исследования получены количественные и качественные характеристики префиксов как в целом по словарю, так и отдельно по частям речи. Рассмотрены приставки одиночные, а также в сочетаниях. Исследованы СГ-структуры префиксов и начальных слогов корней.

Показано, что при очевидной самостоятельности и универсальности значения префиксов сохраняется тенденция к созданию оптимальных ударно-слоговых структур, определяющих форму производного слова.

## Л и т е р а т у р а

1. КОСТОМАРОВ В.Г. и др. Итоги и проблемы подготовки научных и научно-педагогических кадров по языкознанию в 1981-1985 гг. // Вопросы языкознания. - 1986. - №5. - С.3-14.
2. Уровни языка в речевой деятельности: К проблеме лингвистического обеспечения автоматического распознавания речи // Под ред. Л.В.Бондарко. - Л., 1986. - 260 с. (ЛГУ)
3. ЕФРЕМОВА Т.Ф. Из наблюдений над структурой современного русского языка на уровне морфов // Семантические и фонологические проблемы прикладной лингвистики. - М., 1968. - С.45-55. (МГУ)
4. ГЕРДТ А.С. Русская морфология и машинный фонд русского языка // Вопросы языкознания. - 1986. - №6. - С.90-97.

5. WORTH D.S., KOZAK A.S., JOHNSON D.B. Russian derivational dictionary.- New York, 1970.- 747 p.

6. КУЗНЕЦОВА А.И., ЛАВРЕНОВА О.А. О существовании корреляции между продуктивностью и употребительностью аффиксов в русском языке // Исследования по структурной и прикладной лингвистике.- М., 1975. - С.83-99. (МГУ)

7. ЕФРЕМОВА Т.Ф. Несколько замечаний о структурных типах единиц морфного уровня // Семантические и фонологические проблемы прикладной лингвистики. - М., 1968. - С.56-73. (МГУ)

8. БОНДАРКО Л.В. Слоговая структура речи и дифференциальные признаки фонем: Автореф. дис. докт. филолог. наук: 10.02.01. - Л., 1969. - 29 с.

9. ЮДИНА Л.С. Автоматизация транскрипции и некоторые количественные характеристики русской речи: Автореф. дис. канд. филолог. наук: 10.02.01 - Л., 1974. - 23 с.

10. САЛОМАТИНА Н.В., ЮДИНА Л.С. Статистические характеристики консонантных начал слов // Анализ символьных последовательностей. - Новосибирск, 1985. - Вып. 113: Вычислительные системы. - С. 58-83.

Поступила в ред.-изд.отд.

II февраля 1987 года