

# Квазикорпусное изучение словарной продуктивности лексико-синтаксических разрядов в русском языке по словарю С. И. Ожегова<sup>1</sup>

## «Quasi-corpus» investigation of lexical productivity of non-trivial basic diatheses of Russian with special regard to S. I. Ozhegov's dictionary of russian

**Крылов С. А.** (krylov-58@mail.ru)

Институт востоковедения РАН, Институт системного анализа РАН

«Квазикорпусная» лингвистика исследует не только первичные, но и вторичные источники информации о языке (напр., грамматики и словари). Изучена статистика лексико-синтаксических разрядов (подклассов частей речи, отличающихся лексико-синтаксическими пометами) в словаре С. И. Ожегова 1989 г.

В отличие от «корпусной» лингвистики в узком смысле слова, «квазикорпусная» лингвистика (составляющая часть корпусной лингвистики в широком смысле слова) допускает исследование статистики употребления некоторых единиц в составе источников не только первичного характера, но также и вторичного (метаязыкового) характера, в частности, в составе текстов реально существующих грамматик и словарей.

Одной из таких задач стало изучение статистики грамматических помет в словаре С. И. Ожегова 1989 г. (далее СО-1989). В рамках данной задачи можно выделить более узкую — подсчёт частотности лексико-синтаксических помет (далее ЛСП).

На основе помет СО-1989 был создан инвентарь условных маркеров, включающих наряду с исходной ЛСП некоторые обобщённые символы со значением «синтаксических форм», а именно:

A (беспредложный вин. п.)

C (содержание мысли)

D (беспредложный дат. п.)

E (элативная форма)

F (лативная форма)

G (беспредложный род. п.)

I (беспредложный твор. п.)

Inf (инфинитив)

L (эссивная форма)

N (нуль)

R («прочие» синтаксические формы)

S (комитативная форма)

T (делиберативная форма)

V (маршрут //транслативная форма).

С помощью интегрированной информационной среды StarLing, созданной С. А. Старостиним<sup>2</sup>, была создана специальная лексическая база данных на основе СО-1989. Количество лексических единиц (словозначений) в этой базе составляет 64346 (для простоты ЛСП фразем и идиом не учитывались).

<sup>1</sup> Настоящее исследование выполнено при поддержке РФНФ (гранты № 08-04-00190а, № 08-04-12126в, № 08-04-12132в, № 07-04-00161а).

<sup>2</sup> См.: Крылов С. А. Стратегии применения интегрированной информационной среды StarLing в корпусной лингвистике и в компьютерной лексикографии // Смирнов И. С. (ред.). *Orientalia et classica*. Труды Института восточных культур и античности. Выпуск XIX. Аспекты компаративистики. 3. М.: РГГУ, 2008. — С. 649–668.

Идентификация лексико-синтаксических разрядов (ЛСР) производилась на основании сочетания «общекатегориальной» пометы (далее ОКП), указывающей на часть речи или на некоторый грамматический подкласс внутри этой части речи, с ЛСП, указывающей на «исходную диатезу», то есть на тот тип соотношения между смысловыми и синтаксическими валентностями данной лексической единицы (далее ЛЕ), который присущ ей в синтаксически немаркированном употреблении<sup>3</sup>.

Диатезы (в других терминах, «модели управления» или «конструкции») бывают исходные и производные (трансформированные). Исходные диатезы (ИД) бывают тривиальные и нетривиальные.

Тривиальные ИД непосредственно выводятся из ОКП, а потому в словарях задаются фигурой умолчания; как правило, это абсолютные (непереходные) ИД. Их словарная продуктивность максимальна (см. в табл. 1 ЛСР №№ 1–3, 6–7, 9, 12, 14, 20–22, 24–25, 28, 38, 41, 43, 47, 49, 52, 68–69, 78, 87, 89, 97–98, 110, 119–120, 122–123, 125, 149, 157, 178–181, 187 и т. п.).

Содержательно каждая ЛСП отражает некоторую нетривиальную ИД. Нетривиальные ИД, зафиксированные в ЛСП, имеют меньшую словарную продуктивность (см. в табл. 1 ЛСР №№ 4–5, 8, 10–11, 13, 15–19, 23, 26–27, 29–30 и т. п.).

Всего в СО-1989 зафиксировано 980 ЛСР (в ранговом словаре они упорядочены от № 1 до № 980).

Столбец А указывает на ранги; столбец В — на относительную продуктивность ЛСР (т. е. результат деления абсолютной продуктивности на 6,4346); столбец С — на продуктивность ЛСР в абсолютном измерении (количество ЛЕ данного ЛСР). Столбец D указывает на условные маркеры ЛСР.

**Табл. 1.** Список 246 ЛСР, имеющих относительную продуктивность выше 0.31 (а абсолютную — выше 2).

А ранг	В доля на 10 000	С абсолют- ное число	Д обозначение ЛСР
1	4469,90	28762	сущ.
2	1606,32	10336	глагол.
3	1526,75	9824	прил.
4	730,58	4701	глагол. А [что]
5	444,63	2861	глагол. А [кого-что]
6	144,53	930	нареч.
7	115,47	743	первая часть сложных слов
8	100,39	646	глагол. А [кого (что)]
9	47,87	308	приставка

А ранг	В доля на 10 000	С абсолют- ное число	Д обозначение ЛСР
10	42,43	273	глагол. А [что] // G [чего]
11	30,30	195	сущ. G [чего]
12	27,97	180	частица
13	27,35	176	глагол. I [чем]
14	20,51	132	в знач. сказ.
15	20,05	129	глагол. S [с кем]
16	17,72	114	глагол. F [на кого-что]
17	16,32	105	глагол. F [во что]
18	15,23	98	глагол. D [кому]
19	13,99	90	предлог с род. п. G [кого-чего]
20	13,21	85	вводн. сл.
21	12,59	81	мест. нареч.
22	11,50	74	союз
23	10,72	69	глагол. S [с кем-чем]
24	10,57	68	неизм.
25	10,26	66	прил. притяж.
26	9,79	63	глагол. I [кем-чем]
27	9,64	62	глагол. D [кому-чему]
28	9,64	62	межд.
29	9,32	60	глагол. G [чего]
30	9,17	59	глагол. R [Inf]
31	8,70	56	глагол. А [что] & F [во что]
32	8,55	55	глагол. F [к кому-чему]
33	8,55	55	предлог с род. п. G [чего]
34	6,99	45	глагол. L [в чём]
35	6,53	42	глагол. E [от кого-чего]
36	6,37	41	глагол. N [Impers.!] ]
37	6,22	40	глагол. F [на что]
38	6,22	40	числит. колич.
39	5,91	38	глагол. А [что] & I [чем]
40	5,75	37	глагол. А [что] & D [кому]
41	5,59	36	мест. неопр.
42	5,44	35	глагол. E [от чего]
43	5,44	35	сущ. обращение
44	5,28	34	глагол. G [кого-чего]
45	4,97	32	глагол. F [на кого (что)]
46	4,82	31	глагол. S [с чем]
47	4,82	31	мест. определит.
48	4,51	29	сущ. G [кого-чего]
49	4,20	27	числит. собир.
50	3,89	25	глагол. F [к чему]
51	3,73	24	глагол. F [до чего]
52	3,73	24	межд. звукоподр.
53	3,57	23	глагол. F [в кого-что]
54	3,26	21	в знач. сказ. R [Inf]
55	3,26	21	глагол. С [ ]
56	3,26	21	предлог с тв. п. I [кем-чем]
57	3,11	20	глагол. А [что] // R [Inf]

<sup>3</sup> Синтаксически немаркированное употребление глагола обычно представляет собой его употребление в составе так называемой «ядерной конструкции» (см.: Апресян Ю. Д. Экспериментальное исследование семантики русского глагола. М., Наука, 1967, с. 48).

А ранг	В доля на 10 000	С абсолют- ное число	Д обозначение ЛСР
58	3,11	20	предлог с вин. п. А [кого-что]
59	3,11	20	сущ. G [кого]
60	2,95	19	глагол. F [к кому]
61	2,95	19	глагол. L [над кем-чем]
62	2,95	19	глагол. N [Impers.!] & D [кому]
63	2,80	18	глагол. А [кого-что] & I [чем]
64	2,80	18	глагол. F [за кого-что]
65	2,80	18	предлог с вин. и предл. п. А [] // L []
66	2,64	17	мест. нареч. и союзн. сл.
67	2,64	17	предлог с дат. п. D []
68	2,64	17	числит. порядк.; сущ.
69	2,49	16	мест. указат.
70	2,49	16	предлог с вин. п. А [что]
71	2,33	15	глагол. А [что] & S [с чем]
72	2,33	15	глагол. А [что] // T [о чём]
73	2,18	14	в знач. сказ. D [кому]
74	2,18	14	глагол. А [кого-что] & F [во что]
75	2,18	14	глагол. G [кого (чего)]
76	2,18	14	глагол. I [кем]
77	2,18	14	сущ. Adj
78	2,18	14	числит. порядк.; прил.; сущ.
79	2,02	13	глагол. А [кого (что)] & I [чем]
80	1,86	12	глагол. N [Impers.!] & A [кого-что]
81	1,86	12	предлог с тв. п. I [чем]
82	1,71	11	глагол. А [кого-что] // G [чего]
83	1,71	11	глагол. А [кого (что)] & F [во что]
84	1,71	11	глагол. А [что] & F [на кого-что]
85	1,71	11	глагол. D [чему]
86	1,71	11	глагол. L [за кем-чем]
87	1,71	11	мест. притяж.
88	1,71	11	предлог с дат. п. D [кому-чему]
89	1,71	11	союз и частица
90	1,55	10	глагол. А [кого-что] & F [во что]
91	1,55	10	глагол. А [кого-что] & D [кому]
92	1,55	10	глагол. F [за что]

А ранг	В доля на 10 000	С абсолют- ное число	Д обозначение ЛСР
93	1,55	10	глагол. S [с чем] // R [Inf]
94	1,55	10	нареч. и предлог с род. п.
95	1,40	9	глагол. А [что] // F [через что]
96	1,40	9	глагол. T [о ком-чём]
97	1,40	9	мест. отриц.
98	1,24	8	возглас
99	1,24	8	глагол. А [кого]
100	1,24	8	глагол. А [кого (что)] & F [во что]
101	1,24	8	глагол. А [кого (что)] & I [чем]
102	1,24	8	глагол. А [что] & F [на что]
103	1,24	8	глагол. А [что] // I [чем]
104	1,24	8	глагол. E [из чего]
105	1,24	8	глагол. F [до кого-чего]
106	1,24	8	глагол. G [кого-чего] // R [Inf]
107	1,24	8	глагол. L [в ком-чём]
108	1,24	8	глагол. N [Impers.!] & A [кого (что)]
109	1,24	8	глагол. N [Impers.!] & R [Inf]
110	1,24	8	мест. вопросит. и союзн. сл.
111	1,24	8	предлог с вин. и тв. п. А [кого-что] // I [кем-чем]
112	1,09	7	глагол. А [кого-что] & D [кому]
113	1,09	7	глагол. А [что] & F [к чему]
114	1,09	7	глагол. E [с кого-чего]
115	1,09	7	глагол. E [с чего]
116	1,09	7	глагол. F [на что] // R [Inf]
117	1,09	7	глагол. G [чего] // R [Inf]
118	1,09	7	глагол. N [Impers.!] & D [кому] & R [Inf]
119	1,09	7	нареч.; сущ.
120	1,09	7	пов.
121	1,09	7	сущ. F [на что]
122	1,09	7	числит. неопр.-колич.
123	1,09	7	числит. порядк.
124	0,93	6	в знач. сказ. D [кому] & R [Inf]
125	0,93	6	вводн. сл. и частица
126	0,93	6	глагол. А [кого-что] & D [кому-чему]
127	0,93	6	глагол. А [кого-что] & S [с кем-чем]

А ранг	В доля на 10 000	С абсолют- ное число	Д обозначение ЛСР
128	0,93	6	глагол. А [что] & Е [из чего]
129	0,93	6	глагол. D [кому] & R [Inf]
130	0,93	6	глагол. N [Impers. ++]
131	0,93	6	нареч. G [кого-чего]
132	0,93	6	предлог с вин. п. А[]
133	0,93	6	предлог с предл. п. L [ком-чём]
134	0,93	6	прил. D [кому-чему]
135	0,78	5	в знач. сказ. G [кого-чего]
136	0,78	5	в знач. сказ. G [кого-чего] // R [Inf]
137	0,78	5	глагол. А [кого-что] // (А [что] & F [на кого-что])
138	0,78	5	глагол. А [кого-что] // I [чем]
139	0,78	5	глагол. А [кого-что] // R [Inf]
140	0,78	5	глагол. А [кого-что] // устар. G [кого-чего]
141	0,78	5	глагол. А [кого (что)] & L [в чём]
142	0,78	5	глагол. А [кого (что)] F [к чему]
143	0,78	5	глагол. А [что] & D [кому-чему]
144	0,78	5	глагол. А [что] & F [на кого (что)]
145	0,78	5	глагол. А [что] // G [кого-чего]
146	0,78	5	глагол. А [что] // T [о ком-чём]
147	0,78	5	глагол. Е [от кого]
148	0,78	5	глагол. F [к чему] // R [Inf]
149	0,78	5	нареч. и в знач. сказ.
150	0,78	5	предлог с предл. п. L []
151	0,78	5	предлог с предл. п. L [чём]
152	0,78	5	предлог с род. п. G [кого]
153	0,78	5	прил. F [к чему]
154	0,78	5	прил. F [на что] // R [Inf]
155	0,78	5	прил. G [чего] // I [кем-чем]
156	0,78	5	прил. I [чем]
157	0,78	5	сравн.
158	0,78	5	сущ. (обычно в обращении).
159	0,62	4	глагол. (А [кого-что] // I [чем]) & V [по чему]

А ранг	В доля на 10 000	С абсолют- ное число	Д обозначение ЛСР
160	0,62	4	глагол. А [кого-что] & F [со словами «куда», «куда-то», «некуда»]
161	0,62	4	глагол. А [кого-что] & F [до чего]
162	0,62	4	глагол. А [кого-что] // T [о ком-чём]
163	0,62	4	глагол. А [кого (что)] & (F [к чему] // R [Inf])
164	0,62	4	глагол. А [кого (что)] & (F [на что] // R [Inf])
165	0,62	4	глагол. А [кого (что)] & E [от кого-чего]
166	0,62	4	глагол. А [кого (что)] & R [Inf]
167	0,62	4	глагол. А [кого (что)] & F [на что]
168	0,62	4	глагол. А [что] & F [в кого-что]
169	0,62	6	глагол. D [кому] & R [Inf]
170	0,62	4	глагол. Е [с кого]
171	0,62	4	глагол. F [за что] // R [Inf]
172	0,62	4	глагол. F [на кого]
173	0,62	4	глагол. L [за кем-чем] // (устар.) А [что]
174	0,62	4	глагол. N [Impers.!] & А [что]
175	0,62	4	глагол. N [Impers.!] & D [кому]
176	0,62	4	глагол. T [о чём]
177	0,62	4	глагол. V [по кому-чему]
178	0,62	4	звукоподр.
179	0,62	4	мест. личн. 1 л. мн. ч.
180	0,62	4	мест. личн. 3 л.
181	0,62	4	нареч. и союз
182	0,62	4	предлог с дат. п. D [чему]
183	0,62	4	предлог с род. п. G [кого (чего)]
184	0,62	4	прил. F [к кому-чему]
185	0,62	4	прил. S [с чем]
186	0,62	4	прил. в знач. сказ. R [Inf]
187	0,62	4	сущ.; нареч.; неизм.
188	0,62	4	сущ. (обычно в обращении)
189	0,62	4	сущ. G [чего] // Adj [какая]
190	0,62	4	сущ. L [в чём]
191	0,62	4	частица С глаголами прошедшего времени образует сослагательное наклонение

А ранг	В доля на 10 000	С абсолют- ное число	Д обозначение ЛСР
192	0,62	4	частица и в знач. сказ.
193	0,47	3	в знач. сказ. D [кому-чему]
194	0,47	3	в знач. сказ. D [кому] & R [Inf]
195	0,47	3	в знач. сказ. и нареч.
196	0,47	3	восклицание
197	0,47	3	глагол. A [кого-что] & D [чему]
198	0,47	3	глагол. A [кого-что] & F [к кому-чему]
199	0,47	3	глагол. A [кого-что] & E [от кого-чего]
200	0,47	3	глагол. A [кого-что] & E [от чего]
201	0,47	3	глагол. A [кого-что] & F [за кого-что]
202	0,47	3	глагол. A [кого-что] & F [на кого-что]
203	0,47	3	глагол. A [кого-что] & S [с кем-чем]
204	0,47	3	глагол. A [кого (что)] & F [на что]
205	0,47	3	глагол. A [кого (что)] & T [о чём]
206	0,47	3	глагол. A [кого (что)] & E [от кого-чего]
207	0,47	3	глагол. A [кого (что)] & E [от чего]
208	0,47	3	глагол. A [кого (что)] & F [за что]
209	0,47	3	глагол. A [кого (что)] & I [кем]
210	0,47	3	глагол. A [что] & E [от чего]
211	0,47	3	глагол. A [что] & F [подо что]
212	0,47	3	глагол. A [что] & I [кем-чем]
213	0,47	3	глагол. A [что] & S [с кем]
214	0,47	3	глагол. A [что] // A [что] & S [с чем]
215	0,47	3	глагол. A [что] // C [с союзом «что»]
216	0,47	3	глагол. A [что] // F [на что]
217	0,47	3	глагол. A [что] // V [че- рез что]
218	0,47	3	глагол. D [кому] & ( A [что] // R [Inf])
219	0,47	3	глагол. D [кому] & A [что]
220	0,47	3	глагол. D [кому] // F [к кому]

А ранг	В доля на 10 000	С абсолют- ное число	Д обозначение ЛСР
221	0,47	3	глагол. D [кому] // L [перед кем]
222	0,47	3	глагол. E [от чего] // R [Inf]
223	0,47	3	глагол. F [во что] // R [Inf]
224	0,47	3	глагол. F [подо что]
225	0,47	3	глагол. G [кого-чего] // A [что]
226	0,47	3	глагол. L [на чём]
227	0,47	3	глагол. L [перед кем-чем]
228	0,47	3	глагол. N [Impers.!] & A [кого-что]
229	0,47	3	глагол. S [с кем-чем] & L [в чём]
230	0,47	3	глагол. S [с кем (чем)]
231	0,47	3	глагол. T [о ком-чём] // C[с союзом «что»]
232	0,47	3	мест. вопросит.
233	0,47	3	мест. нареч., с последующим отрицанием.
234	0,47	3	мест. указат. и определит.
235	0,47	3	нареч. G [чего]
236	0,47	3	нареч. S [с кем-чем]
237	0,47	3	предлог с вин. и твор. п. A [что] // I [чем]
238	0,47	3	предлог с вин. п. F [кого-что]
239	0,47	3	предлог с вин. п. F [что]
240	0,47	3	прил. S [с кем-чем]
241	0,47	3	сущ.; неизм.
242	0,47	3	сущ. D [кому]
243	0,47	3	сущ. F [на кого-что]
244	0,47	3	сущ. Poss [чей] // G [кого]
245	0,47	3	частица при вопросе
246	0,47	3	числит.

Остальные 734 ЛСР (№№ 247–980) имеют ещё меньшую словарную продуктивность (относительную менее 0,47, а абсолютную менее 3).

В числе уникальных ЛЕ (входящих в одноэлементные ЛСР, то есть с относительной словарной продуктивностью 0,31, а абсолютной 1) оказались представители разных частей речи и грамматических разрядов, а именно: 321 — глаголы; 55 — существительные; 40 — частицы; 35 — прилагательные; 28 — наречия; 23 — союзы; 21 — неглагольные предикативы (помета «в знач. сказ.»); 21 — местоимения; 15 — местоименные наречия; 13 — предлоги; 12 —

междометия; 5 — вводные слова; 4 — числительные; 2 — связки; 2 — сравн.; 1 — союзное слово; 1 — субст. мн.; 1 — нескл.; 1 — возглас; 1 — восклицание; 1 — зват.; 1 — окрик; 1 — ответ; 1 — отклик; 1 — уверение; 1 — форма пов.; 1 — форма 3 л. ед.ч.; 1 — форма 3 л. мн.ч.; 1 — форма род. п. в знач. притяж.

Классификация ЛЕ может основываться не на простом пронумерованном перечне ЛСР, а на определённом ограниченном наборе синтаксических дифференциальных признаков (СДП), характеризующих каждый ЛСР. Такой набор складывается из двух частей.

Во-первых, это (1) ОКП (включая указание (1А) на «часть речи» данной ЛЕ и (1Б) на тот «внутричлестеречный» разряд этой ЛЕ, который зафиксирован в СО-1989).

Во-вторых, это (2) ЛСП (характеристика ИД), складывающаяся из (2А) пометы о безличности/личности данной ЛЕ; (2Б) РП (характеристика типа управления). В РП входят следующие характеристики (СДП):

(2Б.1) синтаксическая абсолютность/релятивность (отсутствие/наличие дополнений);

(2Б.2) общее количество дополнений;

(2Б.3) указание на форму выражения каждого дополнения, то есть на синтетический падеж (А, G, D, I), на предложно-падежную форму (F, L, E, S, T), на инфинитив (Inf), на прилагательное (Adj: *какой*); в том числе притяжательное — *Росс: чей*), на «пропозитив» (изъяснительное придаточное предложение) (С: с союзом *что*). Среди предложно-падежных форм выделяются эссивные (L: *на кого-что, в кого-что, под кого-что, за кого-что, перед кем-чем*), лативные (F: *на кого-что, в кого-что, к кому-чему, под кого-что, до кого-чего*), элативные (E: *от кого-чего, из кого-чего, с кого-чего,* ), транслативные (V: *через кого-что, по кому-чему*), комитативные (S: *с кем-чем*), делиберативные (T: *о ком-чём*)<sup>4</sup>.

<sup>4</sup> Категориальные признаки, лежащие в основе классификации синтаксических форм, заимствованы из разных источников, в том числе из номенклатуры пространственных значений в классификации А. Е. Кибрика (Кибрик А. Е. К типологии пространственных значений // Звегинцев В. А. (отв. ред.). Язык и человек. М.: Изд-во Моск. ун-та, 1970).

Некоторые из выделяемых классов имеют аналоги в классификации синтаксических форм Г. А. Золотовой (см.: Золотова Г. А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса. Изд. 3-е. М.: УРСС, с. 430–432), однако отличаются от синтаксических форм Г. А. Золотовой тем, что используемые в настоящей модели синтаксические формы выделяются не на семантической, а на сугубо формальной (морфологической) основе. Такой подход позволяет описывать систему предложно-падежных форм, являющихся выразителями тех или иных семантических ролей, безотносительно к тому, в какой степени решён вопрос об инвентаризации самих этих семантических ролей. Дело в том, что наиболее естественная номенклатура синтаксических форм, базирующаяся на их основных (главных, первичных) синтаксических функциях, по определению имеет право не принимать во внимание их вторичные функ-

(2Б.4) указание на семантическую одушевлённость (= «одухотворённость») дополнений. Этот признак принимает 4 значения:

(а) отчётливая «одухотворённость»: А (*кого*) и т. п.

(б) тенденция к «одухотворённости»: А (*кого (что)*) и т. п.

(в) нейтральность (диффузность, немаркированность): А (*кого-что*) и т. п.;

(г) отчётливая «неодухотворённость»: А (*что*) и т. п.

Из таблицы 1 (в её полном варианте) нетрудно получить данные о словарной продуктивности любого из этих СДП и любого пучка нескольких СДП (При этом между СДП, входящими в один пучок, можно задавать отношения конъюнкции и дизъюнкции; любой СДП может быть подвергнут логическому отрицанию).

Отдельной интересной задачей было бы соотнесение рангов словарной продуктивности ЛСР с рангами их текстовой частотности, что требует уже применения собственно корпусных методов.

Ранги текстовой частотности ЛСР можно измерять несколькими способами с разной степенью погрешности.

«Корпусные» методы измерения базируются на обследовании достаточно представительного корпуса предварительно размеченных текстов, где каждому слову приписан некоторый ЛСР (разумеется, для измерения частотности реальных диатез каждой текстовой словоформе (каждому вхождению каждой словоформы в текст) следовало бы приписать ту реальную диатезу, в которой эта словоформа употреблена).

Достижение большей или меньшей точности в таком измерении напрямую зависит от того, производится ли разметка (1) вручную или (2) автоматически (ручная разметка обеспечивает более высокую точность измерения).

Автоматическая разметка может базироваться (2а) на моносемизированном корпусе (со снятой омонимией и со снятой полисемией), (2б) на частично моносемизированном корпусе (только со снятой омонимией, но без снятой полисемии) или (2в) на максимально «сыром» (не моносемизированном) корпусе (в котором не снята даже омонимия).

Степень точности (а значит, и адекватности) измерения падает в направлении от способа (1) (где она максимальна) к способу (2в) (где она ми-

ни. Применительно к системам синтетически выражаемых падежей такой «формалистический» подход является практически общепринятым, так что в данном случае речь идёт всего лишь о том, чтобы последовательно применять его к системам падежей, выраженных аналитически — сочетаниями слов-реляторов, или «прилогов» (в частности, в русском языке — предлогов) с формами синтетически выражаемых падежей (см.: Яхонтов С. Е. Классы глаголов и падежное оформление актантов // Храковский В. С. (отв. ред.). Проблемы теории грамматического залога. Л.: Наука, ЛО, 1978, с. 102–108).

нимальна); однако коэффициент трудоёмкости разметки, напротив, очевидным образом оказывается минимальной при выборе способа (2в) и достигает максимума при выборе способа (1).

Трудозатраты на разметку равны результату произведения коэффициента трудоёмкости разметки на объём исследуемого корпуса.

Было бы ошибочным полагать, что всякое измерение частотности ЛСР, обладающее менее чем 100%-ной адекватностью, является бесполезным. Ведь степень адекватности такого измерения находится в прямой зависимости от степени моносемичности разметки, а последняя зависит не только от качества самой разметки, но и от качества исследуемого корпуса: чем меньше доля омонимичных и полисемичных словоформ в исследуемом корпусе, тем более адекватным окажется и само измерение текстовой частотности ЛСР.

Измерения системной мощности (относительной словарной продуктивности) и текстовой мощности (относительной встречаемости в корпусе) ЛСР имеет разнообразные приложения. Результаты этих измерений могут быть использованы в нижеследующих сферах.

- (1) Сопоставительная характеристика языков и количественная типология.

Так, во французском языке (как показал В. Г. Гак) встречаемость транзитивных диатез в речи сравнительно с интранзитивными оказывается значительно выше, чем в русском языке.

Синтаксическая типология (ср. различие эргативных vs. аккумулятивных языков, *быть*-языков vs. *иметь*-языков и т. п.) также оказывается неполной без измерения встречаемости разных конструкций (и выделения шкалы постепенных переходов между синтаксическими типами языков).

- (2) В преподавании языка (как родного, так и неродного), необходима иерархизация ЛСР по количественному параметру ядерности/периферийности: в более кратких описаниях языка можно ограничиваться лишь более ядерными типами ЛСР, а в более полных — включать также более периферийные типы ЛСР.
- (3) При разработке моделей АОТ упорядочение ЛСР по степени их словарной и текстовой мощности предполагает построение синтаксических анализаторов, выстроенных в цепочку последовательных аппроксимаций (где начальная модель учитывает лишь основные типы конструкций, а каждая последующая модель охватывает постоянно растущее число более маргинальных конструкций), что повышает степень адекватности модели.