
ОПЫТ КОРПУСНОГО АНАЛИЗА ЧЕШСКОЙ ОБИХОДНО-РАЗГОВОРНОЙ ЛЕКСИКИ В СОПОСТАВИТЕЛЬНОМ ПЛАНЕ

А.И. Изотов

Кафедра славянской филологии

Филологический факультет

Московский государственный университет им. М.В. Ломоносова

Ленинские горы, ГСП-1, 1-й корпус гум. ф-в, Москва, Россия, 119991

На материале входящего в состав Чешского национального корпуса 1,3-миллиардного под-корпуса современных письменных текстов SYN определяются наиболее употребительные слова (165 лексем с документированной частотой употребления свыше 1000 контекстов) обиходно-разговорного чешского языка, имеющие однословные литературные эквиваленты.

Ключевые слова: чешский язык как иностранный, литературный язык, обиходно-разговорный язык, корпусные исследования, лексикография.

1. Смена научной парадигмы и связанный с ней пересмотр фундаментальных положений прежней теории порой приводят к необходимости переоценки прочно установленных, казалось бы, фактов и утверждений. Весьма показательной в этом плане оказалась ситуация с интерпретацией феномена так называемой чешской диглоссии и связанными с ней базовыми для богемистики понятиями «литературного чешского языка» и «обиходно-разговорного чешского языка». Полвека назад — в эпоху глобального перехода от системно-структурного принципа лингвистического описания к принципу функциональному — эмоциональная дискуссия ведущих чешских лингвистов о статусе «обиходно-разговорного чешского языка» докатилась до станиц советского журнала «Вопросы языкознания» [9; 1]. Не менее эмоциональная дискуссия на ту же тему с участием не менее именитых богемистов разгорелась сегодня, когда все большее значение в лингвистическом исследовании приобретают языковые корпуса, подробнее см. [8].

1.1. Появление в последние годы крупных (миллиарды словоупотреблений) электронных корпусов и связанное с этим беспрецедентное расширение эмпирической базы исследований привело к необходимости пересмотра многих традиционных взглядов на то или иное языковое явление, которые ранее базировались во многом на интуиции исследователя [2]. При этом речь идет не только о возможности надежно установить особенности функционирования ранее недостаточно документированных явлениях языковой периферии [3; 5]. Обращение к корпусу позволяет придать двумерному грамматическому описанию объем за счет введения третьего измерения — статистической характеристики описываемого [4].

1.2. В качестве одного из следствий «корпусной революции» в богемистике [6] стала определенная тенденция к отказу от дихотомии «литературный чешский» — «обиходно-разговорный чешский», господствовавшей в трудах исследователей прошлого века, в пользу дихотомии «письменный чешский» — «устный чешский», которая явственно прослеживается как в современной чешской академической литературе [12], так и в новых учебниках и учебных пособиях по чешскому языку для иностранцев [10].

Впрочем, развитие корпусных технологий может оказаться полезным и для того, кто остается в рамках классической терминологии. Авторы написанной еще в «докорпусную эпоху» монографии [13] предложили лингвистической обществу пробный вариант словаря наиболее важной, по их мнению, чешской обиходно-разговорной лексики (около 1000 единиц), снабдив каждую подобную лексическую единицу литературным эквивалентом (либо толкованием), а также пометой: *běž* — лексическая единица, общая и для обиходно-разговорного, и для литературного чешского языка, *ob* — лексическая единица обиходно-разговорного чешского языка, не маркированная в качестве стилистически сниженной, *n* — лексическая единица обиходно-разговорного чешского языка, маркированная как стилистически сниженная, *n.sl.* — лексическая единица, характерная для нелитературной речи солдат, ремесленников и т.д., *v.sl.* — лексическая единица, характерная для нелитературной речи врачей, инженеров, студентов и т.д., *expr* экспрессивная лексическая единица, *pejor* пейоративная лексическая единица, *zast* устаревающая лексическая единица, *zhrub* грубая лексическая единица.

2. Мы попытались выявить употребительность представленных в данном словаре лексем, а также их литературных эквивалентов в текстах входящего в состав Чешского национального корпуса набора подкорпусов современных письменных текстов SYN общим объемом 1,3 миллиарда токенов [14].

2.1. Поиск однословных лексических единиц осуществлялся с помощью простых запросов типа [lemma="ajznboňák"], находящего контексты употребления заданной леммы (в данном случае леммы *ajznboňák* ‘железнодорожник’) во всех ее словоформах. Ниже представлены некоторые из найденных контекстов.

Jsem <ajznboňák> od dvou let.

Na dvorku se objeví "<ajznboňáci>" v uniformách z počátku století.

Tmavomodrý svět je kusem světa <ajznboňáků>.

2.2. Для поиска неоднословных лексических эквивалентов или же толкований приходилось использовать более сложные запросы, представляющие собой поиск по той или иной комбинации леммы (lemma), словоформы (word) и грамматической матрицы (tag) в пределах одного предложения (within <s id=".*"/>). Например, для поиска контекстов употребления лексического толкования *být potrestán za něco* ‘быть наказанным за что-либо’ использовался запрос ([lemma="být"] [(lemma="potrestat") & (tag="Vs.*")] [word="za"]) | ([lemma="být"] [word="za"]) []{0,3} [(lemma="potrestat") & (tag="Vs.*")] within <s id=".*"/>. Вот некоторые из найденных контекстов:

<Byl za to vzápětí potrestán> a všichni se o potrestání mají dovědět možná jen proto, aby se také začali bát o život.

Spravedlnost vyžaduje, aby seňora <nebyla potrestána za> něco, co neudělala.

Kluci chtěli vědět, jestli <budou za svoje minulé hříchy potrestáni>, jestli na ně jejich holka počká.

2.3. Тем самым определение частотности употребления входящих в однословные соответствия «лексема обиходно-разговорного чешского языка» — «лексема литературного чешского языка с той же семантикой» было наиболее простым

и в то же время наиболее полезным для целей практического преподавания чешского языка как иностранного, так как оно позволяет установить иерархию приоритетов при знакомстве обучаемого с лексикой обиходно-разговорного чешского языка. О том, почему обучение обиходно-разговорному чешскому языку следует начинать с фонетики и лексики, а не с морфологии, см. [11. С. 82].

2.4. Впрочем, и в случае с однословными соответствиями могли возникать определенные трудности.

2.4.1. Во-первых, задавая поиск по лексеме (поисковый запрос типа [lemma="šéf"]), мы вынуждены доверять основанной на стохастическом анализе автоматической разметке корпусного материала, которая может ошибаться. Поэтому у нас нет статистических данных по тем приведенным в [13. С. 139—169] словам, которые были идентифицированы программой автоматической разметки не в качестве самостоятельных лексем, а в качестве фонетических (графических) вариантов соответствующих слов литературного чешского языка: *dybu* (лит. *kdyby*) 'если бы', *dýl* (лит. *déle*) 'дольше', *dyž* (лит. *když*) 'когда', *esli* (лит. *jestliže*) 'если', *eště* (лит. *ještě*) 'еще', *chleba* (лит. *chléb*) 'хлеб', *lip* (лит. *lépe*) 'лучше', *matrace* (лит. *matrace*) 'матрас', *pámbu* (лит. *pánbůh*) 'Господь Бог', *zejtra* (лит. *zítra*) 'завтра'.

Впрочем, подобных слов было не так уж много. Поиск по словоформе был бы более точен, однако он неизмеримо более трудоемок, так как предполагает поиск всех теоретически возможных словоформ заданной лексемы с последующим суммированием результатов, например поиск [word="[Šš]éf"] + поиск [word="[Šš]éfa"] + поиск [word="[Šš]éfu"] + поиск [word="[Šš]éfovi"] + поиск всех оставшихся падежных форм, которых в чешском языке еще больше, чем в русском.

2.4.2. Во-вторых, проблему представляют омонимы, для различения которых требуется визуальный контроль контекстов их употребления, особенно тех из них, которые отличаются высокой частотностью употребления. Например, внимательно просмотреть на экране 104 контекста, обнаруженных с помощью запроса [lemma="fknkout"], чтобы убедиться, что в 102 случаях из них речь идет именно о значении *utéct* 'убежать, ускользнуть', можно достаточно быстро, благо все найденные контексты достаточно однозначны. А вот на то, чтобы определить, в каком из 4271 контекста, обнаруженном по запросу [lemma="flek"], речь идет о значении *místo* '[рабочее] место', а в каком — о значении *skvrna* 'пятно' (оба значения отмечены в [13. С. 146]), может не хватить целого вечера.

2.4.3. В аналогичной ситуации мы оказываемся во всех тех случаях, когда в [13. С. 139—169] приводятся стоящие рядом омонимы, ср. некоторые примеры, снабженные нашим переводом на русский язык:

foch	obor ['специальность']	ob	
foch	příhrádka ['полка']	ob	
...
lágr	ložisko ['месторождение']	n.sl.	
lágr	tábor ['лагерь']	ob	
...
pingl	číšník ['официант']	ob	expr
pingl	ranec ['ранец']	ob.	

2.4.4. К рассмотренным выше примерам примыкают и, к сожалению, несравненно более многочисленные случаи, когда в качестве литературного эквивалента обиходно-разговорной лексемы в [13. С. 139—169] приводятся (через запятую либо точку с запятой) два или более слова, например:

akorát	zrovna [‘как раз’], přesně [‘точно’]; jen [‘только’]	ob	
...
brajgl	hluk [‘шум’], nerořádek [‘беспорядок’]	ob	rejor
...
fofr	shon [‘суматоха’], spěch [‘спешка’]	ob	expr
...
pusa	ústa [‘рот’], polibek [‘поцелуй’]	běž.	

Допуская, что в чешской языковой картине *суматоха* (shon) может восприниматься как нечто тождественное или по крайней мере весьма близкое *спешке* (spěch), а *шум* (hluk) — *беспорядку* (nerořádek), мы тем не менее подозреваем, что по крайней мере в части случаев (например, в последнем из приведенных выше), речь идет все же отнюдь не о синонимах. Однако в любом случае статистическая обработка подобных примеров невозможна без предварительного визуального контроля каждого контекста их употребления.

2.4.5. Наконец, следует отметить случаи, когда в [13. С. 139—169] речь идет не об обиходно-разговорной лексеме, а о лексеме литературной, однако имеющей в обиходно-разговорном языке дополнительное (нелитературное) значение, например:

balík [‘пакет’]	hlupák [‘глупец’]	běž	rejor
...
bažant [‘фазан’]	nováček [‘новичок’]	n. sl.	
...
beran [‘баран’]	tvrdohlavec [‘упрямец’]	ob	expr.

3. В результате исключения названных выше требующих дополнительной обработки соответствий мы получили таблицу, начало которой (с числом документированных употреблений в корпусе SYN больше 1000) выглядит следующим образом (табл. 1).

Таблица 1

Однословные соответствия

Слово	Идиом	Стиль	Число найденных контекстов	Литературное лексическое соответствие	Число найденных контекстов	Лексическое соответствие в русском
gól	běž	—	400 702	branka	450 042	гол
šéf	běž	—	359 917	vedoucí	372 890	заведующий
jes(t)li	běž	—	289 763	jestliže	109 327	если
taky	běž	—	189 554	také	2 278 504	тоже
pořád	ob	—	181 478	stále	697 389	постоянно
fanoušek	běž	—	155 431	přivřezec	8 207	сторонник
kluk	běž	—	151 781	chlapec	138 205	мальчик
konto	běž	—	82 695	účet	124 996	счет

Слово	Идиом	Стиль	Число найденных контекстов	Литературное лексическое соответствие	Число найденных контекстов	Лексическое соответствие в русском
muzika	<i>běž</i>	—	57 737	hudba	291 000	музыка
holka	<i>běž</i>	—	52 764	děvče	55 809	девушка
jo / jó	<i>ob</i>	—	48 807	ano	214 221	да
dneska	<i>běž</i>	—	45 686	dnes	11 646 010	сегодня
muzikant	<i>běž</i>	—	45 577	hudebník	46 140	музыкант
chlap	<i>běž</i>	—	45 246	muž	1 013 293	мужчина
táta	<i>běž</i>	—	43 707	otec	238 950	отец
koukat (se)	<i>běž</i>	—	31 831	dívat se	61 268	смотреть
máma	<i>běž</i>	—	31 231	matka	219 242	мать
kanonýr	<i>n.sl.</i>	—	30 217	dělostřelec	1 304	артиллерист
legrace	<i>běž</i>	—	27 223	zábava	90 742	развлечение
radši	<i>běž</i>	—	26 115	raději	127 801	лучше
fotka	<i>běž</i>	—	22 765	fotografie	271 923	фотография
kasa	<i>ob</i>	—	20 266	pokladna	66 250	касса
řadit	<i>běž</i>	—	20 100	stranit	2 673	симпатизировать
teda	<i>běž</i>	—	17 187	tedy	806 817	следовательно
policajt	<i>ob</i>	—	13 771	policista	512 709	полицейский
bál	<i>běž</i>	—	13 354	ples	46 454	бал
barák	<i>běž</i>	<i>expr</i>	12 804	dům	953 081	дом
bankrot	<i>běž</i>	—	12 758	úpadek	13 060	крах
kantor	<i>běž</i>	<i>expr</i>	11 907	učitel	153 553	учитель
cedule	<i>běž</i>	—	11 554	vývěska	2 836	вывеска
řítit	<i>běž</i>	—	11 407	fotografovat	13 968	фотографировать
Pražák	<i>běž</i>	—	10 925	Pražan	29 374	пражанин
kafe	<i>běž</i>	—	9 990	káva	47 713	кофе
štvát	<i>běž</i>	—	9 412	rozčilovat	14 593	раздражать
huba	<i>ob</i>	<i>zhrub</i>	9 194	ústa	60 866	рот
naštvát	<i>běž</i>	—	9 118	rozčlít	6 310	разозлить
fabrika	<i>ob</i>	—	9 069	továrna	60 304	фабрика
sranda	<i>ob</i>	—	8 975	zábava	90 742	развлечение
krám	<i>běž</i>	—	8 917	obchod	325 185	магазин
děcko	<i>běž</i>	—	8 861	dítě	1 415 924	ребенок
praštit	<i>běž</i>	—	8 743	uhodit	7 864	ударить
prachy	<i>ob</i>	<i>expr</i>	8 685	peníze	815 548	деньги
brácha	<i>ob</i>	<i>expr</i>	8 444	bratr	150 313	брат
průšvih	<i>běž</i>	—	8 108	nepříjemnost	9 031	неприятность
brečet	<i>běž</i>	—	7 963	plakat	23 406	плакать
fáma	<i>běž</i>	—	7 779	pověst	61 008	молва
pasovat někam	<i>ob</i>	—	7 757	hodit se	20 390	подходить [для чего]
řacka	<i>ob</i>	—	7 741	pohlavek	1 377	оплеуха
kriminál	<i>ob</i>	—	7 657	vězení	134 751	заключение
lump	<i>běž</i>	—	7 285	darebák	5 557	бездельник
špitál	<i>běž</i>	—	7 224	nemocnice	485 266	больница
blbec	<i>běž</i>	<i>pejor</i>	7 050	pitomec	3 280	дурак
šéfovat	<i>běž</i>	—	7 001	řídit	140 698	управлять
ahoj	<i>běž</i>	—	6 842	nazdar	2 201	привет!
túra	<i>n. sl.</i>	—	6 415	cesta	741 732	путешествие
šuplík	<i>ob</i>	—	6 374	zásuvka	10 940	выдвижной ящик
kytka	<i>běž</i>	—	6 109	kytice	14 641	букет
beztak	<i>běž</i>	—	6 089	beztoho	450	и без того
švec	<i>běž</i>	—	5 462	obuvník	1 047	сапожник
pitomý	<i>běž</i>	<i>pejor</i>	5 434	hloupý	25 526	глупый
metrák	<i>běž</i>	—	4 788	cent	10 993	центнер
kámoš	<i>ob</i>	<i>expr</i>	4 769	kamarád	116 395	приятель
halt / holt	<i>ob</i>	—	4 680	zkrátka	52 273	просто-напросто
marod	<i>běž</i>	—	4 507	nemocný	118 011	больной

Продолжение

Слово	Идиом	Стиль	Число найденных контекстов	Литературное лексическое соответствие	Число найденных контекстов	Лексическое соответствие в русском
kurva	<i>ob</i>	<i>zhrub</i>	4 137	prostitutka	17 759	проститутка
polda	<i>běž</i>	<i>expr</i>	4 001	policista	512 709	полицейский
furt	<i>ob</i>	—	3 867	stále	697 389	постоянно
chmatač	<i>ob</i>	<i>expr</i>	3 848	zloděj	191 928	вор
hajzl	<i>n</i>	—	3 543	záchod	17 747	туалет
rande	<i>běž</i>	—	3 541	schůzka	88 940	встреча
hovno	<i>ob</i>	<i>zhrub</i>	3 388	lejno	846	дерьмо
flaška	<i>ob</i>	—	3 307	láhev	78 241	бутылка
buřt	<i>ob</i>	—	3 299	vuřt	1 462	сарделька
pitomec	<i>běž</i>	<i>pejor</i>	3 280	hloupý	25 526	дурак
kecat	<i>ob</i>	<i>pejor</i>	3 279	žvanit	1 696	болтать
potvora	<i>běž</i>	<i>pejor</i>	3 239	mizera	2 958	негодяй
strejda	<i>běž</i>	<i>expr</i>	3 226	strýc	16 011	дядя
hecovat	<i>běž</i>	—	3 183	dráždit	6 942	дразнить
cpát co kam	<i>běž</i>	—	3 147	mačkat	4 133	давить
marodit	<i>běž</i>	—	3 114	nemocen být	920	болеть
pošťák / počták	<i>ob</i>	—	3 068	listonoš	1 287	почтальон
mrazák	<i>běž</i>	—	3 019	mrazička	1 540	морозилка
kvartál	<i>běž</i>	—	2 968	čtvrtletí	57 041	квартал
gauner	<i>ob</i>	<i>pejor</i>	2 924	darebák	5 557	бездельник
drb	<i>ob</i>	<i>pejor</i>	2 907	pomluva	10 014	сплетня
děvka	<i>běž</i>	<i>pejor</i>	2 888	prostitutka	17 759	проститутка
šenkýř	<i>ob</i>	<i>zast</i>	2 879	hostinský	10 602	трактирщик
šokovat	<i>běž</i>	—	2 834	pohoršovat	2 081	возмущать
kumšt	<i>běž</i>	<i>expr</i>	2 785	umění	265 715	искусство
lamentovat	<i>běž</i>	—	2 784	nařikat	7 133	сетовать
profík	<i>běž</i>	—	2 735	profesionál	33 662	профессионал
šikanovat	<i>běž</i>	—	2 709	pronásledovat	18 940	преследовать
garsonka / garzonka	<i>běž</i>	—	2 705	garsoniéra	1 341	студио
vepředu	<i>běž</i>	—	2 658	vpředu	12 347	впереди
plichta	<i>ob</i>	—	2 588	remíza	79 372	ничья
šidit	<i>běž</i>	—	2 534	ubližovat	5 590	обманывать
špunt	<i>ob</i>	—	2 492	zátka	3 023	затычка
fešák	<i>běž</i>	—	2 457	elegán	1 451	франт
kuráž	<i>ob</i>	—	2 445	odvaha	39 671	отвага
nátura	<i>ob</i>	—	2 399	povaha	45 938	характер
figl	<i>ob</i>	—	2 340	trik	17 687	уловка
špek	<i>běž</i>	—	2 330	slanina	6 382	шпик
míň	<i>běž</i>	—	2 325	méně	67 430	меньше
šichta	<i>ob</i>	—	2 276	směna	25 162	смена
rachejtle	<i>ob</i>	<i>expr</i>	2 213	raketa	37 727	ракета
teďka	<i>ob</i>	—	2 208	teď	604 771	сейчас
flinta	<i>běž</i>	—	2 205	puška	16 124	ружье
perón	<i>běž</i>	—	2 138	nástupiště	7 433	перрон
parchant	<i>ob</i>	<i>pejor</i>	2 120	darebák	5 557	бездельник
kýbl	<i>ob</i>	—	2 036	kbelík	4 318	ведро
bachař	<i>n.sl.</i>	—	2 013	dozorce	9 942	охранник
šlapka	<i>běž</i>	<i>zast</i>	1 958	prostitutka	17 759	проститутка
ksicht	<i>n</i>	—	1 957	obličej	74 850	лицо
pech	<i>ob</i>	<i>expr</i>	1 936	smůla	41 199	невезение
plac	<i>ob</i>	—	1 929	místo	1 783 165	место
polívka	<i>běž</i>	—	1 867	polévka	23 494	суп
grázl	<i>ob</i>	<i>pejor</i>	1 860	darebák	5 557	бездельник
syčák	<i>běž</i>	<i>pejor</i>	1 792	ničema	2 019	негодяй
kumštýř	<i>běž</i>	<i>expr</i>	1 748	umělec	121 726	деятель искусства
basta	<i>ob</i>	—	1 708	došt!	1 985	хватит!

Слово	Идиом	Стиль	Число найденных контекстов	Литературное лексическое соответствие	Число найденных контекстов	Лексическое соответствие в русском
mrcha	<i>běž</i>	<i>zhrub</i>	1 678	mizera	2 958	негодяй
šesták / šesták	<i>ob</i>	<i>zast</i>	1 672	dvacetihaléf	194	монета в 20 геллеров
fotr	<i>n</i>	—	1 631	otec	238 950	отец
lejstro	<i>běž</i>	<i>pejor</i>	1 614	listina	50 297	документ
kravál	<i>běž</i>	<i>pejor</i>	1 590	rámus	3 593	гам
placatý	<i>ob</i>	—	1 550	plochý	17 348	плоский
čumák	<i>běž</i>	—	1 546	čenich	1 259	морда
šenk	<i>ob</i>	—	1 528	výčep	4 280	распивочная
pysk	<i>běž</i>	—	1 501	reť	29 104	губа
mrňavý	<i>běž</i>	<i>expr</i>	1 489	malý	1 018 346	щуплый
fuška	<i>běž</i>	<i>expr</i>	1 479	dřina	11 298	тяжелый труд
foť	<i>ob</i>	—	1 471	lesník	8 803	лесник
šiml	<i>běž</i>	—	1 456	bělouš	1 495	лошадь светлой масти
strejček	<i>běž</i>	—	1 434	strýc	16 011	дядя
čuník	<i>běž</i>	<i>expr</i>	1 427	prase	22 445	свинья
čumět	<i>ob</i>	<i>zhrub</i>	1 422	zírat	13 351	смотреть
gympl	<i>v.sl.</i>	—	1 398	gymnázium	99 946	гимназия
mukl	<i>n.sl.</i>	—	1 371	vězeň	60 594	заключенный
mus	<i>ob</i>	—	1 370	nutnost	48 316	необходимость
pendrek	<i>běž</i>	—	1 343	obušek	4 198	дубинка
veselka	<i>běž</i>	<i>zast</i>	1 337	svatba	55 860	свадьба
putyka	<i>ob</i>	<i>pejor</i>	1 314	hosпода	71 437	пивная
biják	<i>ob</i>	—	1 302	kino	284 124	кинотеатр
paraple	<i>běž</i>	<i>zast</i>	1 296	deštník	7 518	зонтик
fajfka	<i>ob</i>	—	1 290	dýmka	8 115	курительная трубка
prkotina	<i>běž</i>	—	1 289	maličkost	14 209	мелочь
kšeftovat	<i>ob</i>	—	1 286	obchodovat	25 727	торговать
sesypat se	<i>ob</i>	—	1 249	zhrouřit se	6 475	развалиться
piáno	<i>běž</i>	—	1 244	klavír	24 517	рояль
prekérní	<i>běž</i>	—	1 240	povážlivý	1 293	опасный
srab	<i>ob</i>	<i>pejor</i>	1 227	zbabělec	3 296	трус
mlíko	<i>běž</i>	—	1 209	mléko	56 717	молоко
čouhat	<i>běž</i>	—	1 203	vyčnívat	4 917	выступать
mord	<i>n</i>	—	1 183	vražda	115 236	убийство
šmejd	<i>ob</i>	<i>pejor</i>	1 176	brak	1 899	дешевка
hasrman	<i>ob</i>	—	1 174	vodník	6 820	водяной
ruksak	<i>ob</i>	—	1 144	batoh	15 306	рюкзак
culit se	<i>běž</i>	<i>expr</i>	1 136	usmívat se	36 077	улыбаться
Moravák	<i>běž</i>	—	1 135	Moravan	12 332	мораванин
vytahovat se	<i>běž</i>	—	1 131	chlubit se	8 433	хвастаться
šutr	<i>ob</i>	—	1 101	kámen	106 188	камень
rošťák	<i>běž</i>	—	1 061	uličník	1 638	хулиган
štrúdl	<i>běž</i>	—	1 055	závin	1 398	рулет
kapku	<i>běž</i>	—	1 028	trochu	315 246	немного
penál	<i>běž</i>	—	1 025	pouzdro	9 509	футляр

3.1. Данная таблица однословных соответствий, которую мы планируем представить целиком в одном из ближайших выпусков продолжающегося сборника «Язык, сознание, коммуникация» [15], позволяет определить не только абсолютную, но и относительную частоту употреблений рассматриваемых лексем. Например, если мы примем общее количество зафиксированных употреблений той или

иной обиходно-разговорной лексемы (А) и ее литературного коррелята (В) за 100%, то относительная частотность данной обиходно-разговорной лексемы будет равняться $A/(A + B) \times 100\%$.

3.2. В соответствии с данной таблицей и данной формулой расчета (А — сумма документированных контекстов употребления всех лексем из первой колонки; В — сумма документированных контекстов употребления всех их литературных коррелятов) средняя относительная частотность лексемы из первой колонки составляла приблизительно 5,2%.

При этом средняя относительная частотность лексических единиц, характерных исключительно для обиходно-разговорного чешского языка и не маркированных в качестве стилистически сниженных (снабженных пометой *ob*), **в три раза ниже** относительной частотности лексических единиц (снабженных пометой *běž*): речь идет о 2,5% и о 7,5% соответственно. Если же мы будем учитывать также и лексические единицы обиходно-разговорного чешского языка, маркированные как стилистически сниженные (помета *ob* + помета *n*), то эта относительная частотность будет еще ниже — 2,1%.

Что же касается стилистических помет, то весьма высокой оказалась относительная частотность лексических единиц, маркированных как *грубые* — 9,1%, тогда как у слов, маркированных как *нейоративные*, *экспрессивные* или *устаревшие*, относительная частотность оказалась существенно ниже средней — 1,79%, 0,95% и 0,84% соответственно.

4. Полученный нами частотный словарь соответствий чешской обиходно-разговорной и литературной лексики не претендует на то, чтобы быть исчерпывающим уже в силу того, что на это не претендует список лексем, изначально предложенный П. Сгаллом и И. Гронекком. Тем не менее употребительность в современных чешских письменных текстах (а во многих случаях — весьма высокая употребительность) тех лексических единиц обиходно-разговорного чешского языка, которые были рассмотрены в данной статье, следует считать эмпирически доказанной, что безусловно должно быть учтено как минимум в лексикологии современного чешского языка, стилистике, практической лексикографии и лингводидактике.

ЛИТЕРАТУРА

- [1] *Белич Я., Гавранек Б., Едличка А., Травничек Ф.* К вопросу об «обиходно-разговорном» чешском языке и его отношении к литературному чешскому языку // Вопросы языкознания. — 1961. — № 1. — С. 44—51.
- [2] *Изотов А.И.* Опыт корпусного анализа функционально-семантической категории императивности в современном чешском языке // Вестник Московского университета. Серия 9: Филология. — 2007. - № 3. — С. 47—63.
- [3] *Изотов А.И.* Чешский национальный корпус и аналитический императив: опыт корпусного анализа малоупотребительных и маргинальных языковых единиц // Вестник Оренбургского государственного университета. — 2007. — № 2. — С. 4—11.

- [4] *Изотов А.И.* Императивность как прагмалингвистический феномен: на материале чешского языка. — М.: Книжный дом «ЛИБРОКОМ», 2008.
- [5] *Изотов А.И.* Опыт корпусного анализа чешских директивных и директивно-комиссивных перформативных глаголов // Вестник Российского университета дружбы народов. Серия «Лингвистика». — 2012. — № 3. — С. 99—107.
- [6] *Изотов А.И.* Корпусная революция: от «искусства» к «науке» // Филологические науки: Вопросы теории и практики. — 2013. — № 4—1. — С. 68—71.
- [7] *Изотов А.И.* Чешское обиходно-разговорное койне vs русское просторечье // Вестник Российского университета дружбы народов. Серия «Лингвистика». — 2013. — № 1. — С. 52—60.
- [8] *Изотов А.И.* Обиходно-разговорный чешский язык — химера или реальность? // Альманах современной науки и образования. — 2013. — № 12. — С. 69—72.
- [9] *Сгалл П.* Обиходно-разговорный чешский язык // Вопросы языкознания. — 1960. — № 2. — С. 11—20.
- [10] *Cvejnová J.* *Česky, prosím II.* Učebnice češtiny pro cizince. — Praha: Karolinum, 2012.
- [11] *Hrdlička M.* Kapitoly o češtině jako cizím jazyce. — Plzeň: Západočeská univerzita v Plzni, 2010.
- [12] *Mluvená čeština: hledání funkčního rozpětí* / Eds. S. Čmejrková, J. Hoffmannová. — Praha: Academia, 2011.
- [13] *Sgall P., Hronek J.* Čeština bez příkras. — Praha: H&H, 1992.
- [14] URL: <http://ucnk.ff.cuni.cz>.
- [15] URL: http://www.philol.msu.ru/~slavphil/books/jsk_index.html.

COMPARATIVE CORPUS ANALYSIS OF EVERYDAY SPOKEN CZECH LANGUAGE AND LITERARY LANGUAGE

A.I. Izotov

Department of Slavic Philology
Faculty of Philology
Lomonosov Moscow State University
Leninskie Gory, 1 Humanities Building, Moscow, 119192

In this paper the lexis of everyday spoken Czech is compared with literary Czech lexis. The data is taken from the material which is a part of the Czech National Corpus of 1.3 billion contemporary written texts and identifies the most common words. A list of 165 pairs of everyday spoken Czech lexemes and literary Czech lexemes with the same lexical meaning has been established.

Key words: Czech as a foreign language, Literary Czech, everyday spoken language, corpus analysis, lexicography.

BIBLIOGRAPHY

- [1] *Belich I., Gavranek B., Edlichka A., Travnichek Ph.* К вопросу об «обиходно-разговорном» чешском языке и его отношении к литературному чешскому языку [To the problem of every day spoken Czech language and its correlation with literary language] // Вопросы языкознания. — 1961. — № 1. — P. 44—51.

- [2] *Izotov A.I.* Opit korpusnogo analiza funktsionalno-semanticheskoy kategorii imperativnosti v sovremennom cheshskom yazyke [Comparative corpus analysis of functional-semantic imperative category in modern Czech language] // *Vestnik Moskovskogo universiteta. Seria 9: Philologia*. — 2007. — № 3. — P. 47—63.
- [3] *Izotov A.I.* Cheshskiy natsionalniy corpus I analiticheskiy imperativ: opit korpusnogo analiza malopotrebitelnikh I marginalnikh yazykovikh edinit [Czech national corpus and analytical imperative: comparative corpus analysis of rare and marginal words] // *Vestnik Orenburgskogo gosudarstvennogo universiteta*. — 2007. — № 2. — С. 4—11.
- [4] *Izotov A.I.* Imperativnost kak pragmalinguisticheskii fenomen: na materiale cheskogo yazyka [Imperative as a pragmalinguistics phenomena: based on Czech language]. — М.: «LIBROKOM», 2008.
- [5] *Izotov A.I.* Opit korpusnogo analiza cheshskikh direktivnikh i direktivno-komissionnikh performativnikh glagolov [Comparative corpus analysis of directive and commission performative verbs] // *Vestnik Rossiiskogo universiteta druzbi narodov. Series: Linguistics*. — 2012. — № 3. — P. 99—107.
- [6] *Izotov A.I.* Korpusnaya revolutsia: ot “iskusstva” k “nayke” [Corpus-bases revolution: from “art” to “science”] // *Philologicheskie nauki: voprosi teorii I praktiki*. — 2013. — № 4—1. — P. 68—71.
- [7] *Izotov A.I.* Cheshckoe obihodno-razgovornoe koine vs russkoe prostorechie [Every day spoken Czech Koine vs Russian popular language] // *Vestnik Rossiiskogo universiteta druzbi narodov. Series: Linguistics*. — 2013. — № 1. — С. 52—60.
- [8] *Izotov A.I.* Obihodno-razgovornii cheshskii — khimera ili realnost? [Every day spoken Czech — chimera or reality?] // *Almanakh sovremennoy nayki i obrazovania*. — 2013. — № 12. — С. 69—72.
- [9] *Sgall P.* Obihodno-razgovornii cheshskii [Every day spoken Czech] // *Voprosi yazikoznania*. — 1960. — № 2. — С. 11—20.