

**М.Н. Михайлов**  
(Тампере)

## **Корпус текстов как инструмент лингвистического мониторинга**

### **1. Общие замечания**

Исследования диахронических изменений, происходящих в языке за относительно короткие промежутки времени (от десяти до ста лет), практически невыполнимы традиционными методами. Ведь изменения часто оказываются незаметными «невооруженным глазом»: одни слова выходят из употребления, другие — наоборот активизируются; появляются новые значения, изменяется модель управления и т.п. Незначительность таких изменений вовсе не означает их несущественность, ибо капля точит камень. А знания о существующих тенденциях в развитии языка представляются столь же полезными, как прогноз погоды и рейтинги политиков.

Наличие больших размеченных корпусов текстов и возможность быстрого выполнения запросов делает задачу проведения лингвистических мониторинговых исследований более реальной (см., например, Integrum 2006). В этой статье мы постараемся продемонстрировать возможности проведения такого мониторинга на примере слов с семантикой умственной деятельности. Исследуемое семантическое поле является довольно большим и включает в себя и существительные (*голова, мозг, мысль, мышление, сознание, интеллект* и т.д.), и прилагательные (*умный, интеллектуальный, разумный* и т.д.), и наречия (*умно*), и глаголы (*думать, мыслить, размышлять, теоретизировать* и т.д.), а также множество фразеологизмов (*в здравом уме, в полном рассудке, без царя в голове, ума палата* и т.д.). В данной статье будет исследоваться употребление наиболее частотных слов этой группы в художественных текстах. Поскольку в нашем распоряжении имеется двуязычный языковой материал, показалось интересным сравнить процессы, идущие в разных языках, а также межъязыковое взаимодействие через художественный перевод (ср. тж. соображения о полезности привлечения многоязычного материала в Dagan et al. 1991).

Основным источником языкового материала стал русско-финский корпус параллельных художественных текстов «ПарРус», собранный на кафедре перевода русского языка Института современных языков и переводоведения Тамперского университета (Финляндия). В состав корпуса

включены произведения русской художественной литературы XIX-XX вв. и их переводы на финский язык, изданные в 1950-2000 гг. Объем корпуса составляет около 5 млн. словоупотреблений, тексты корпуса выровнены и лемматизированы. Более подробно корпус описан в (Михайлов 2003). Для проведения исследования тексты корпуса были разбиты на три субкорпуса: тексты второй половины XIX в., первой половины XX в. и второй половины XX в.<sup>1</sup>

Для получения аналогичных данных по финскому языку использовался массив современных художественных текстов, входящих в Савонлиннский корпус текстов<sup>2</sup> (тексты 1990-х гг.), и массив финских художественных текстов 1870-х—1920-х гг. (последние получены из текстового архива «Гутенберг»: <http://promo.net/pg/>)<sup>3</sup>.

Количественные данные по экспериментальным текстовым массивам приведены в таблицах 1 и 2.

**Таблица 1. Корпус «ПарРус»: количество словоупотреблений**

	Русские тексты	Финские переводы
с 1950	1074362	1013467
1900-1950	513901	484086
1850-1900	697977	691955
<b>Всего:</b>	<b>2286240</b>	<b>2189508</b>

**Таблица 2. Корпус финской прозы: количество словоупотреблений**

1870-1920	554891
1990-2000	1390590
<b>Всего:</b>	<b>1945481</b>

## 2. Наблюдения над частотностью употребления

Это мини-исследование началось с получения странных данных по частотности существительных, обозначающих умственную способность: *ум, разум, рассудок, мозг*. Оказалось, что частотность этих слов в современной русской литературе заметно ниже, чем в литературе второй половины 19 века (см. табл. 3): абсолютное количество контекстов, полученных из классической литературы, чуть меньше, чем в современной, при том, что субкорпус современной литературы почти в два раза больше. Только по слову *мозг* (по вполне понятным причинам) наблюдается рост, по всем остальным словам частотность снижается, причем включение данных по первой половине 20 века показывает, что снижение происходит планомерно<sup>4</sup>. Проверка этих данных по Национальному корпусу русского языка (НКРЯ, [www.ruscorpora.ru](http://www.ruscorpora.ru)) в целом наши на-

блюдения тоже как будто подтверждает (см. табл. 4), хотя здесь также следует соблюдать некоторую осторожность в суждениях, поскольку объем субкорпуса современной художественной литературы в НКРЯ еще сильнее отличается по размерам от субкорпусов литературы первой половины XX века и второй половины XIX века, чем в «ПарРус», втроекратно превосходя их по количеству словоупотреблений.

Полученные из корпусов данные представляются по меньшей мере странными: можно согласиться, что слово *мозг* в языке XIX века относилось скорее к научному дискурсу, можно поверить, что слово *рассудок* несколько архаично и на пути в пассив (хотя в МАС у этого слова никаких стилистических помет нет). Но почему частотность слова *ум* так сильно упала за сто с небольшим лет? Не является ли это косвенным свидетельством «перехода от слов к делу» или какого-нибудь «горя от ума»?

МАС выделяет в данных лексемах следующие значения: 1) ‘способность логически мыслить, понимать окружающий мир’ (*мозги, ум, разум, рассудок*), 2) ‘орган мышления’ (*мозг*), 3) ‘человек как носитель этой способности’ (*мозг, ум*), 4) ‘нормальное состояние сознания человека, способность адекватно реагировать на обстановку’ (*рассудок, разум, ум*), 5) ‘здравый смысл» как противоположность эмоциям’ (*разум, рассудок, ум*).

Таблица 3. Частотность существительных группы УМ в «ПарРус»<sup>5</sup>

	1850-1900		1900-1950		после 1950	
мозг	13	0,02	52	0,1	85	0,08
разум	33	0,05	13	0,03	23	0,02
рассудок	47	0,07	4	0,01	12	0,01
ум	203	0,29	115	0,22	179	0,17
<b>Всего:</b>	<b>296</b>	<b>0,43</b>	<b>184</b>	<b>0,36</b>	<b>299</b>	<b>0,28</b>

Таблица 4. Частотность существительных группы УМ в Национальном корпусе русского языка

	1850-1900		1900-1950		после 1950	
мозг	326	0,03	1033	0,08	3266	0,09
разум	597	0,05	530	0,04	1393	0,04
рассудок	267	0,02	123	0,01	372	0,01
ум	4083	0,36	2696	0,22	7302	0,21
<b>Всего:</b>	<b>5273</b>	<b>0,46</b>	<b>4382</b>	<b>0,35</b>	<b>12333</b>	<b>0,35</b>

Несмотря на такую кажущуюся близость значений, нельзя утверждать, что данные слова являются синонимами. Рассмотрим четыре примера, в

которых слова *мозги*, *ум*, *рассудок* и *разум* употреблены в значении ‘способность логически мыслить’:

- (1) *Что знает рассудок? Рассудок знает только то, что успел узнать (иного, пожалуй, и никогда не узнает; это хоть и не утешение, но отчего же этого и не высказать?) а натура человеческая действует вся целиком, всем, что в ней есть, сознательно и бессознательно, и хоть врет, да живет* (Ф. Достоевский, «Записки из подполья»).
- (2) *Да и зачем тебе ум, если у тебя есть совесть и сверх того еще и вкус? Совесть и вкус — это уж так много, что мозги становятся прямо излишними* (Вен. Ерофеев, «Москва-Петушки»).
- (3) *... Смычков вытащил из футляра контрабас. Минуту казалось ему, что он, уступая футляр, профанирует святое искусство, но колебание было непродолжительным. Красавица легла в футляр и свернулась калачиком, а он затянул ремни и стал радоваться, что природа одарила его таким умом* (А. Чехов, «Роман с контрабасом»).
- (4) *Насколько я заметила, каждое водворение этой молодой власти проходит через несколько этапов. В начале это торжество разума, критический дух, борьба с предрассудками* (Б. Пастернак, «Доктор Живаго»).

В примере (1) слово *рассудок* употреблено в значении ‘способность анализировать полученную информацию’. Пожалуй, в этом контексте слово свободно от каких-либо семантических приращений. Однако уже в примере (2) речь идет не просто о способности мыслить, но о способности мыслить неординарно, показывать высокий интеллект. В этом примере слова *ум* и *мозги* употреблены как синонимы. В примере (3) имеется в виду практический ум, сообразительность, умение находить оригинальные решения. А в примере (4) способность мыслить представлена как мышление вообще, безотносительно к субъекту.

Таким образом, в разных контекстах может идти речь о довольно разных способах проявления мыслительной способности. Поэтому перед тем, как начинать делать какие бы то ни было обобщения, стоит посмотреть контексты употребления этих слов в корпусе. Вот, например, некоторые примеры употребления слова *ум*:

- (5) *Я без ума, без памяти, без конца люблю тебя* (Б. Пастернак, «Доктор Живаго»).

- (6) *Он медленно сходил с ума* (Б. Пастернак, «Доктор Живаго»).
- (7) *Мы научим уму разуму* (Б. Пастернак, «Доктор Живаго»).
- (8) *Такое лицо у него было, когда он решал в уме навигационную задачу* (А. и Б. Стругацкие, «Попытка к бегству»).
- (9) *Какой это был дерзкий ум!* (А. и Б. Стругацкие, «Попытка к бегству»).
- (10) *Я думаю, что равнодушие есть лучшее из состояний человеческого ума* (Ю. Олеша, «Зависть»).
- (11) *А ишо вот чего... — Мать зашептала: — Возьми да в уме помолись* (В. Шукшин, «Материнское сердце»).
- (12) *Где у вас директор? самое сильное, что пришло Сашке на ум* (В. Шукшин, «Обида»).

Из примеров (5)-(12) видно, что слово *ум* далеко не всегда употребляется в значении «способность к абстрактной мыслительной деятельности». Довольно большой процент употреблений приходится на различные устойчивые обороты: *сойти / сходить с ума, прийти на ум, быть без ума, делать что-либо в уме* и т.п. Поэтому нельзя сказать, что частотность слова *ум* в тексте указывает на описание в тексте умственных процессов.

Итак, колебания в частотности слова могут быть связаны не только с семантикой, но и с грамматикой, с тем, какие синтаксические конструкции употребляются в текстах. Может быть, наблюдаемое нами загадочное падение частотности для данной группы слов компенсируется увеличением частотности синонимичных им слов, относящихся к другим грамматическим категориям? Например, снижается частотность оборота *прийти / придти на ум* и при этом возрастает частотность глаголов *думать / подумать*. Для проверки этой гипотезы потребуется сравнение частотности разных лексем, связанных с мыслительной деятельностью, причем список должен быть по возможности полным.

Кроме лексики, описывающей способность к мыслительной деятельности как объект (*ум, разум, группа УМ*), в том же поле можно выделить лексику, описывающую мыслительную деятельность как признак (*умный, умница, глупый, дурак, далее — группа УМНЫЙ*), и конечно же лексику, описывающую процесс мышления (*думать, мысль, далее — группа ДУМАТЬ*).

### 3. Семантическая группа УМНЫЙ:

#### способность к интеллектуальной деятельности как признак

К этой группе относятся как прилагательные, так и существительные. Все слова оказываются в той или иной мере оценочными, и всю лексику этой группы можно разделить на лексику позитивной (*умный, смысленный*) и негативной (*глупый, идиот*) оценки. Результаты, полученные из корпуса «ПарРус», представлены в таблице 5. Слова с частотностью ниже 0,03 на 1000 словоупотреблений во всех субкорпусах в учет не принимались.

Таблица 5. Частотность слов группы УМНЫЙ в корпусе «ПарРус»

	1850-1900		1900-1950		с 1950	
глупый	210	0,3	68	0,13	119	0,11
дурак	82	0,12	77	0,15	191	0,18
дурацкий	9	0,01	21	0,04	55	0,05
идиот	5	0,01	10	0,02	37	0,03
мудрый	0	0	10	0,02	28	0,03
разумный	19	0,03	9	0,02	34	0,03
сумасшедший	44	0,06	78	0,15	27	0,03
толк	33	0,05	38	0,07	86	0,08
толковый	4	0,01	7	0,01	27	0,03
умный	179	0,26	59	0,11	147	0,14
<b>Всего:</b>	<b>585</b>	<b>0,84</b>	<b>377</b>	<b>0,73</b>	<b>751</b>	<b>0,7</b>

В первой же строке таблицы привлекает внимание падение частотности слова *глупый* с 0,3/1000 до 0,11/1000, а в последней строке видим, что и слово *умный* стало употребляться существенно реже: частота последнего упала с 0,26 на 1000 словоупотреблений у классиков до 0,14 в современной литературе. В целом же, тенденция, замеченная нами в предыдущей группе, сохраняется и здесь, правда падение частотности менее заметное. Но следует все же отметить, что самое заметное снижение частотности наблюдается именно для общеупотребительной лексики.

### 4. Семантическая группа ДУМАТЬ:

#### мыслительная деятельность как таковая

Эта семантическая группа оказалась самой большой и самой спорной по составу. Представляется, что в поле следует включить слова *помнить, понимать* и др., поскольку в противном случае мы неоправданно сузим поле до собственно аналитической деятельности, между тем как группа глаголов *прийти на ум / подумать / понять / вспомнить* воспринимается

как вполне гомогенная. Поэтому мы пытались по возможности учитывать и такую лексику.

Ядро группы составляют глаголы, что автоматически означает существенное удлинение списка за счет видовых пар и префиксальных дериватов. Наибольшую трудность вызывают глаголы типа *думать*, которые образуют стремящиеся к бесконечности ряды типа *подумать*, *выдумать*, *придумать*, *надумать*, *передумать*, *обдумать*, *задумать* и т.д. Подсчет частотности каждого из этих глаголов по отдельности вылился бы в весьма трудо- и времязатое занятие, причем трудно сказать, удалось ли бы учесть все возможные дериваты. По этой причине частотность рассчитывалась для всего гнезда сразу, что, к счастью, оказалось технически нетрудно реализовать. Разумеется, всегда существует вероятность, что у некоторых из анализируемых слов имеются значения, выходящие за рамки данной семантической группы.

Во всех случаях, исследуя семантическую группу, не следует забывать, что значение слова выходит за рамки грамматических категорий. То есть кроме глаголов, мыслительный процесс может обозначаться и с помощью существительных (*память*, *идея* и др.), многие из которых являются именными производными от глаголов (*мыслить* — *мышление*, *мысль*).

И вновь наше предположение о том, что снижение частотности слов *ум* и *разум* может компенсироваться повышением частотности других языковых средств выражения аналогичных значений, не оправдалось. Мы вновь наблюдаем общее снижение относительной частотности. Наибольшее изменение отмечено у слова *мысль* (с 0,9/1000 до 0,39/1000).

**Таблица 6. Частотность слов группы ДУМАТЬ в корпусе «ПарРус»**

	1850-1900		1900-1950		с 1950	
(*)думать(ся) <sup>6</sup>	2106	3,02	993	1,93	2700	2,51
(*)думывать(ся)	98	0,14	40	0,08	139	0,13
идея	39	0,06	40	0,08	112	0,1
мысль	629	0,9	315	0,61	418	0,39
мнение	138	0,2	44	0,09	60	0,06
память	78	0,11	78	0,15	135	0,13
(*)помнить(ся)	773	1,11	395	0,77	1152	1,07
(*)помянуть(ся)	252	0,36	121	0,24	233	0,22
(*)понимать(ся)	692	0,99	318	0,62	1006	0,94
(*)понять	573	0,82	285	0,55	1176	1,09
представить(ся)	171	0,24	84	0,16	223	0,21
представлять(ся)	133	0,19	80	0,16	173	0,16

	1850-1900		1900-1950		с 1950	
раздумья	26	0,04	6	0,01	27	0,03
размышления	14	0,02	16	0,03	21	0,02
размышлять	14	0,02	20	0,04	45	0,04
(*)считать(ся)	313	0,45	166	0,32	538	0,5
<b>Всего:</b>	<b>6049</b>	<b>8,67</b>	<b>3001</b>	<b>5,84</b>	<b>8158</b>	<b>7,59</b>

## 5. Что происходит в финском языке?

Интересным представляется проверить, имеет ли место аналогичная тенденция в другом языке. Для этого был составлен список финских лексем с аналогичным значением и проверена их частотность на контрольном массиве. Покажем для примера данные, полученные по финским существительным, обозначающим способность к мышлению (таблица 7)<sup>7</sup>. Как видно из таблицы 7, налицо снижение частотности как по всей группе, так и по отдельным словам. По финской лексике, обозначающей мыслительную деятельность как признак, существенных изменений выявлено не было, но по лексике, обозначающей мыслительную деятельность, вновь фиксируется падение общей средней относительной частотности с 7,38/1000 до 6,7/1000. Таким образом, как это ни удивительно, но и финский материал демонстрирует похожую тенденцию: частотность лексики, связанной с мыслительной деятельностью, в современной финской художественной литературе оказывается ниже, чем в литературе XIX — начала XX века.

Таблица 7. Частотность финской лексики группы УМ в финской художественной литературе

	1870-1920		Совр. тексты	
aivot 'мозги'	35	0,06	159	0,11
järki 'ум'	78	0,14	151	0,11
mieli 'душа, разум'	1275	2,3	2507	1,8
tolkku 'толк, здравый смысл'	1	0	34	0,02
ymmärrys 'разум'	22	0,04	20	0,01
äly 'ум, интеллект'	7	0,01	39	0,03
<b>Всего:</b>	<b>1418</b>	<b>2,56</b>	<b>2910</b>	<b>2,09</b>

Располагая переводами русских литературных текстов на финский язык, мы не могли не проверить частотность тех же лексем в переводных текстах. Ожидалось, что частотность финских эквивалентов исследуемых русских слов будет более или менее близка к русскому прототипу. Предполагалось также, что в случае отсутствия четких соответствий —



как например *ум* → *mieli, järki, äly* — частотности могут «накладываться» или «вычитаться», но не более того.

**Таблица 8. Частотность финской лексики группы УМ в «ПарРус»**

	1850-1900		1900-1950		с 1950	
aivot 'мозги'	17	0,02	55	0,11	104	0,1
järki 'ум'	122	0,18	90	0,19	142	0,14
mieli 'душа, ум'	370	0,53	419	0,87	1081	1,07
äly 'ум, интеллект'	36	0,05	16	0,03	20	0,02
<b>Всего:</b>	<b>528</b>	<b>0,76</b>	<b>525</b>	<b>1,09</b>	<b>1243</b>	<b>1,23</b>

Тем не менее, нас ожидал очередной сюрприз. Во-первых, частотность слов в переводах оказалась совершенно не похожей на частотность этих слов в аутентичных финских текстах (см. табл. 8). Этого, впрочем, вполне можно было ожидать. Во-вторых, частотность этих слов ничем не напоминает частотность русской лексики той же группы в русских оригинальных текстах корпуса (ср. табл. 3.), что само по себе показалось несколько странным. И наконец, в-третьих, в таблице 8 налицо общий рост частотности слов от 0,76/1000 до 1,23/1000, то есть наблюдается тенденция, обратная зафиксированной в финских и русских аутентичных художественных текстах.

**Таблица 9. Частотность финской лексики со значением мыслительной деятельности как признак в «ПарРус»**

	1850-1900		1900-1950		с 1950	
hölmö 'дурак'	62	0,09	40	0,08	141	0,14
hullu 'сумасшедший'	39	0,06	57	0,12	117	0,12
järjetön 'бессмысленный'	24	0,03	7	0,01	22	0,02
järkevä 'разумный'	39	0,06	27	0,06	67	0,07
tyhmä 'глупый'	49	0,07	24	0,05	87	0,09
typerä 'глупый'	99	0,14	59	0,12	73	0,07
viisas 'мудрый'	84	0,12	48	0,1	188	0,19
älykäs 'смышленный'	43	0,06	16	0,03	26	0,03
<b>Всего:</b>	<b>439</b>	<b>0,63</b>	<b>278</b>	<b>0,57</b>	<b>721</b>	<b>0,73</b>

Частотности слов со значением мыслительной деятельности как признака также показывают тенденцию к небольшому росту (см. табл. 9). Снижение частотности наблюдается лишь у трех слов из восьми.

**Таблица 10. Частотность финской лексики со значением мыслительной деятельности в «ПарРус»**

	1850-1900		1900-1950		с 1950	
ajatella 'думать'	483	0,7	569	1,18	1232	1,22
ajatus 'мысль'	202	0,29	273	0,56	411	0,41
arvella 'предполагать'	46	0,07	18	0,04	97	0,1
idea 'идея'	5	0,01	2	0	46	0,05
keksiä 'придумать'	100	0,14	73	0,15	237	0,23
luulla 'думать'	168	0,24	158	0,33	429	0,42
mieltä 'думать'	101	0,15	65	0,13	345	0,34
muistaa 'помнить'	249	0,36	274	0,57	956	0,94
muisti 'память'	91	0,13	118	0,24	379	0,37
pohtia 'думать'	26	0,04	32	0,07	83	0,08
tuumia 'думать'	21	0,03	8	0,02	63	0,06
käsitys 'представление'	36	0,05	33	0,07	59	0,06
käsittää 'представлять'	106	0,15	118	0,24	282	0,28
ymmärtää 'понимать'	325	0,47	313	0,65	1109	1,09
<b>Всего:</b>	<b>1959</b>	<b>2,83</b>	<b>2054</b>	<b>4,25</b>	<b>5728</b>	<b>5,65</b>

И наконец группа слов со значением мыслительной деятельности оказалась существенно больше двух предыдущих (см. табл. 10), но и в этом случае тенденция сохраняется. Более того, общее увеличение относительной частотности оказывается даже более заметным, почти двукратным.

## **6. Обратный параллельный конкорданс как инструмент исследования**

Для выяснения причин этого странного явления обратимся к контекстам. Поскольку ручная проверка всего материала даже для нашего относительно небольшого списка представляется малореалистичной, ограничимся лишь выборочным просмотром параллельных конкордансов на слова, для которых зафиксированы самые значительные различия. Назовем получаемые нами параллельные контексты обратным параллельным конкордансом, поскольку будет выполняться поиск слов и выражений, встретившихся в переводах, и через них будут получены исходные контексты из оригинальных текстов.

Поиск на слово *mieli* 'душа, ум' дал ответ на вопрос, почему у этого слова такая высокая частотность. Дело в том, что многие формы этого слова адвербиализованы (*mielestäni* 'по-моему', *mielelläni* 'с удовольствием'), а другие входят в состав устойчивых оборотов (*samaa / toista mieltä* 'того же / другого мнения', *tekee mieli* 'хочется'), но при этом

нельзя сказать, что какие-либо значения этой леммы выходят за пределы интересующего нас семантического поля. Поэтому во многих случаях в русской части конкорданса в качестве стимулов для *mieli* оказывались слова, отсутствовавшие в наших списках. Тем не менее, среди примеров параллельного конкорданса встретились и такие случаи, когда в русском оригинальном контексте отсутствовали слова исследуемого семантического поля.

- (13) *Но я-то, я, — смотрел на него со злобою и ненавистью, и так продолжалось... несколько лет-с! Злоба моя даже укреплялась и разрасталась с годами* (Ф. Достоевский, «Записки из подполья»).

В переводе (13) на финский язык Э. Адриан использует оборот *minä katselin häntä mieli täynnä kiukkaa ja vihaa* ('я смотрел на него с душой, полной злобы и ненависти').

- (14) *Ну, ладно, — согласился Булдеев. Тут не только что к акцизному, но и к черту депешу пошлешь...* (А. Чехов, «Лошадиная фамилия»).

При переводе (14) Ю. Конкка эксплицирует русский текст: *Tässä tekee mieli lähettää sähkösanoma...* ('Тут хочется дать телеграмму...')

- (15) *Все рассмеялись, так одновременно без слов всем в спальне бросилось в глаза это сходство* (Б. Пастернак, «Доктор Живаго»).

- (16) *...и в свои зауспокойные строки по Анне Ивановне вставит все, что ему к той минуте подвернется, все случайное, что ему подсунет жизнь...* (Б. Пастернак, «Доктор Живаго»).

Переводчик Ю. Конкка использует в качестве эквивалента оборота *броситься в глаза* из примера (15) и для слова *подвернуться* из примера (16) финский оборот *juolahtaa mieleen* ('прийти на ум').

Глагол *ajatella* 'думать' гораздо конкретнее существительного *mieli*, однако в конкордансе на это слово выявляются те же тенденции. Хотя очень во многих случаях этот глагол используется для перевода контекстов с глаголами *думать*, *размышлять* или оборотов *прийти на ум*, *быть на уме* и т.п., встречается достаточно много контекстов, когда финский глагол эксплицирует пресуппозицию русского текста.

- (17) *Кроме того, Юре было приятно, что он опять встретится с Никой Дудоровым, гимназистом, живущим у Воскобойникова* (Б. Пастернак, «Доктор Живаго»).

В финском переводе Ю. Конкка читаем *Jurasta oli mieluista ajatella* ('Юре было приятно думать').

- (18) *В душе у Аннушки все пело от предвкушения того, что она будет завтра рассказывать соседям* (М. Булгаков, «Мастер и Маргарита»).

В переводе У.-Л. Хейно: *Annuška tunsi jo mielessään suloisia väristyksiä ajattellessaan...* ('Аннушка чувствовала уже в душе приятную дрожь, думая...').

Другая особенность примеров конкорданса на глагол *ajatella* состоит в том, что в русском тексте нередко появляются разговорные, книжные или стилистически окрашенные слова, в качестве переводного соответствия которых используется этот нейтральный финский глагол, см. примеры 19-22.

- (19) *Мечик не успел еще сообразить, в чем дело, а уж Бакланов, спешившись, возился у седла* (А. Фадеев, «Разгром»).
- (20) *Надо мыслить, — сурово сказал Остап* (И. Ильф, Е. Петров, «Золотой теленок»).
- (21) *Пятилетка в четыре года, даже в три — вот стимул, который... Возьмите, наконец, даже мою жену. Домашняя хозяйка - и та отдает должное индустриализации* (И. Ильф, Е. Петров, «Золотой теленок»).
- (22) *Опомнись! Зачем тебе ломать свою жизнь с больным и ничим? Вернись к себе!* (М. Булгаков, «Мастер и Маргарита»).

## 7. Интерпретация частотных данных

Различие данных по частотности в аутентичных финских текстах и переводах с русского языка на финский, а также выборочный анализ параллельных конкордансов дает возможность предложить объяснение для обнаруженного нами явления.

По всей видимости, корень различий в частотности лексики в современной литературе и литературе XIX века не в «оглуплении» общества, а в глубоких процессах в развитии языка, стиля и методов художественной литературы. От эксплицитных описаний мышления и состояния духа человека произошел переход к подтексту, потоку сознания и др. приемам литературы XX века. Авторская речь начинает вытесняться прямой и несобственно прямой речью. Эта-то тенденция и фиксируется нашей статистикой. То, что «мысли» на самом деле никуда не исчезли, а

просто ушли в подтекст, показывают частотные данные из переводов, в которых частотность наоборот растет.

Многие переводоведы отмечают в качестве особенностей языка переводов более бедный по сравнению с языком аутентичных текстов лексический и грамматический репертуар, большая степень эксплицитности текста, большая консервативность в языке и стиле (см., например, Newmark 1991; Комиссаров 1992). Поэтому во многих случаях то, что в оригинальном тексте было в подтексте, в переводе оказывается на поверхности. В этом плане переводные тексты оказываются очень полезным источником данных: они служат как бы зеркалом, в котором отражаются языковые особенности оригинала. Причем, как любое зеркало, они не только «отражают» но и «переворачивают» исходный текст.

Итак, в этой статье мы пытались показать, что наличие аннотированных электронных корпусов текстов и соответствующих инструментов для выполнения запросов по ним делает возможным не только изучение текущего состояния языка, но и наблюдение за процессами, в нем происходящими. Вполне возможно, что организация такого мониторинга для самых разных языков — больших и малых — вовсе не является делом далекого будущего.

### Примечания

<sup>1</sup> В корпусе «ПарРус» есть и тексты первой половины XIX века (Пушкин, Лермонтов, Гоголь), однако по объему эта часть корпуса заметно меньше трех других субкорпусов. Поэтому было принято решение не рассматривать эту часть корпуса.

<sup>2</sup> *The Savonlinna Corpus of Translated Finnish*. Savonlinna School of Translation Studies, University of Joensuu, 2001.

<sup>3</sup> Корпуса текстов, на материале которых выполнялось исследование, размещены на сервере кафедры перевода русского языка, доступном через веб-интерфейс по адресу <https://mustikka.uta.fi/corpora>. Доступ к ресурсу ограничен.

<sup>4</sup> Отмечу по ходу дела, что резкое возрастание частотности слова *мозг* в субкорпусе первой половины 20 века в значительной степени связано с включением в корпус повести М. Булгакова «Собачье сердце», в которой, как известно, описывается операция на головном мозге.

<sup>5</sup> В этой и следующих таблицах рядом с абсолютной частотностью указывается относительная на 1000 словоупотреблений, которая является для нашего исследования более существенной, поскольку все текстовые массивы неодинаковы по размерам.

<sup>6</sup> В скобки заключается факультативный элемент строки, звездочка обозначает любую цепочку знаков. Таким образом нотация (\*)думать(ся) означает, что кроме слова *думать* в список включаются также *думаться*, *подумать*, *подуматься*, *задумать*, *задуматься* и т.д. и т.п.

<sup>7</sup> Здесь и далее списки финских соответствий были получены с помощью «Большого русско-финского словаря» М. Куусинена и В. Оллыкяйнен. Далее слова из списка проверялись по корпусам текста, после чего низкочастотные и не встретившиеся ни разу эквиваленты отбрасывались.

**Литература**

CD-Perussanakirja: 1997, Kotimaisten kielten tutkimuskeskus.

Dagan, I.; Itai, A.; and Shwall, U.: 1991, 'Two languages are more informative than one', *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, 130-137.

Integrum: точные методы и гуманитарные науки: 2006, Москва: Летний сад.

Newmark, P.: 1991, 'The Virtues of interference and the vices of translationese', *About Translation*, Clevedon / Philadelphia / Adelaide, 78-87.

Комиссаров, В.Н.: 1992, '«Естественность» художественного перевода', *Литература и перевод: проблемы теории*, Москва, 101-110.

Куусинен М. & Оллыкайнен В.: 1978, *Большой русско-финский словарь*, WSOY, Helsinki.

МАС: 1999, *Словарь русского языка в 4-х т.*, РАН, Ин-т лингвистич. исследований, Москва: Русский язык.

Михайлов, М.Н.: 2003, 'Параллельные корпуса художественных текстов: принципы составления и возможности применения в лингвистических и переводоведческих исследованиях', *Acta Universitatis Tamperensis*, 956, *Acta Electronica Universitatis Tamperensis*, 280, University of Tampere.