

5. Баранов А.Н. Введение в прикладную лингвистику. М.: Издательство ЛКИ, 2007. 360 с.
6. Archiv für gesprochenes Deutsch. <http://agd.ids-mannheim.de> (от 26.05.2009).
7. Fiehler, R., Wagener, P. Die Datenbank Gesprochenes Deutsch (DGD). <http://www.gespraechsforschung-ozs.de/heft2005/px-fiehler.pdf/> (от 15.05.2009).
8. Datenbank gesprochenes Deutsch. <http://dsav-wiss.ids-mannheim.de> (от 26.05.2009).
9. Berliner Wendekorpus. <http://agd.ids-mannheim.de/html/korpora/korpus-bw.shtml> (от 04.10.2009).
10. Dittmar, N. Informationen zu meinen Korpora. http://userpage.fu-berlin.de/~nordit/HP/Ditt_Korpora.html (от 20.09.2005).
11. Dittmar, N., Bredel, U. Die Sprachmauer. Die Verarbeitung der Wende und ihre Folgen in Gesprächen mit Ost- und WestberlinerInnen. Berlin: Weidler Buchverlag, 1999. 208 S.
12. Digitales Wörterbuch der deutschen Sprache DWDS. <http://www.dwds.de> (от 26.05.2009).

*Андрей Борисович КУТУЗОВ —
ст. преподаватель кафедры перевода и переводоведения
Тюменского государственного университета,
кандидат филологических наук
akutuzov72@gmail.com*

УДК 81'255.4

ЧАСТОТНЫЕ ХАРАКТЕРИСТИКИ ЛЕКСИКИ КАК ИНВАРИАНТ ПЕРЕВОДА (НА МАТЕРИАЛЕ ПЕРЕВОДА РОМАНА К. ВОННЕГУТА «БОЙНЯ НОМЕР ПЯТЬ, ИЛИ КРЕСТОВЫЙ ПОХОД ДЕТЕЙ»)*

LEXICAL FREQUENCIES AS TRANSLATIONAL INVARIANCE (BASED ON RUSSIAN TRANSLATION OF K. VONNEGUT'S SLAUGHTERHOUSE-FIVE OR THE CHILDREN'S CRUSADE)

АННОТАЦИЯ. На основе лексико-статистического анализа оригинала и русского перевода романа К. Воннегута Slaughterhouse-Five была выявлена линейная корреляция между динамикой изменения соотношения числа словоформ и словоупотреблений в оригинале и в переводе. Автор делает предположение о том, что эта динамика является инвариантом перевода.

SUMMARY. The article provides lexical statistical analysis of K. Vonnegut's Slaughterhouse-Five and its Russian translation. It is found out that there exists a linear correlation between the speed of types and tokens ratio change in the source and target texts. The author hypothesizes that this dynamics is a translational invariance.

КЛЮЧЕВЫЕ СЛОВА. Переводоведение, корпус, лингвостатистика, словоформа, словоупотребление, переводческий инвариант.

KEY WORDS. Translation studies, corpora, computational linguistics, type, token, translational invariance.

Переводоведение всегда испытывало некоторые трудности с определением своего предмета. Тем более неоднозначным является вопрос о методах этой науки. Как можно объективно и надежно оценить адекватность или эквивалентность перевода? Есть ли в процессе перевода некие четкие зависимости, влия-

* Работа выполнена при поддержке Федерального агентства по науке и инновациям (государственный контракт № 02.740.11.0594).

ющие на его результат? Этими вопросами лингвисты и переводоведы задавались на протяжении всей второй половины прошлого века и продолжают задаваться по сей день.

В конце 1990-х гг. в рамках изучения этой проблематики появились первые попытки объединить переводоведение и корпусную лингвистику с ее точными методами анализа больших текстовых массивов. Квантитативные методы позволяют достаточно подробно и достоверно описать результат перевода и сделать некоторые выводы о его процессе. Обширная литература по этому вопросу приведена, например, в [1].

Неудивительно, что это направление исследований быстро привело переводоведов к вопросу о том, существуют ли некие объективные статистические зависимости между текстом оригинала и перевода. Вопрос этот тесно связан с гипотезой М. Бейкер об объективных отличиях любых переводных текстов от текстов не-переводных, спонтанных [2]. Его решение позволит сравнивать переводы на относительно точной основе, а кроме того, даст новые ключи к пониманию того, как функционирует человеческий мозг в процессе такой сложной деятельности, как межъязыковой перевод. Сохраняются ли в процессе перевода определенные статистические инварианты текста или же перевод носит произвольный характер? Этому вопросу и посвящена настоящая статья.

В качестве переводного корпуса для предварительного исследования этой проблемы был избран роман Курта Воннегута «Slaughterhouse-Five Or The Children's Crusade» (1969) и его перевод на русский язык «Бойня номер пять, или Крестовый поход детей» Риты Райт-Ковалевой (1978). Тексты произведений взяты из электронной библиотеки Максима Мошкова (<http://www.lib.ru>) и с сайта «Книги Курта Воннегута на английском языке» (<http://bg-studio.newmail.ru>). Выявление статистических параметров корпуса производилось при помощи утилиты AntConc (<http://www.antlab.sci.waseda.ac.jp/software.html>). При анализе регистр букв не учитывался, морфологическая обработка не проводилась.

В тексте оригинала романа 6247 словоформ (word types) и 50848 словоупотреблений (word tokens). Таким образом, их соотношение составляет 0,12. 50% словоформ составляют словоформы, употребленные лишь один раз (так называемые *haraх legomena*).

Первые 8 самых частых слов обычны для английского языка — это артикли, местоимения и другие служебные слова. Лишь на 9 месте по частоте употребления (800 словоформ) стоит Billy — имя главного персонажа романа.

При взгляде на частотный словарь произведения хорошо заметно, как уменьшается частота по мере увеличения ранга словоформы. Это следствие из закона Ципфа, который гласит, что частота употребления словоформ в любом тексте на естественном языке обратно пропорциональна их рангу. То есть частых слов всегда мало, а редких слов всегда много [3]. Например, в исследуемом тексте однократно употребленные слова составляют 50% от всего словаря произведения. Именно из закона Ципфа следует низкое количество и высокая частота служебных слов по сравнению со знаменательными.

Мы сравнили между собой каждую из десяти глав романа, чтобы найти объективные статистические закономерности, содержащиеся в произведении. В табл. 1 продемонстрированы соотношения словоформ и словоупотреблений (TTR) в каждой из глав в отдельности:

Таблица 1

Лексикостатистические параметры глав оригинала

Глава	Количество словоформ	Количество словоупотреблений	Соотношение словоформ и словоупотреблений (TTR*)
1	1411	5494	0,26
2	1805	7328	0,25
3	1376	4663	0,3
4	1099	3536	0,31
5	2389	11217	0,21
6	1207	4357	0,28
7	620	1676	0,37
8	1300	4686	0,28
9	1651	6529	0,25
10	523	1240	0,42
Весь роман	6247	50848	0,12 (нормализованное значение 0,29)

Наиболее богатой по лексическому составу является последняя, десятая глава, хотя она самая короткая (соотношение TTR 0,42). То же самое можно сказать о седьмой главе — она лишь немногим длиннее и тоже обладает богатым лексическим запасом (TTR 0,37). Можно сделать вывод, что по мере увеличения длины глав не происходит такого же увеличения количества словоформ — оно растет гораздо медленнее. Это подтверждают и общие параметры романа — TTR по всему роману (0,12) значительно меньше, чем в любой из глав.

Таким образом, мы наблюдаем, что в текстах присутствуют не только статические закономерности, описываемые законом Ципфа, но и другие, динамические закономерности. В частности, распределение частот слов зависит от длины самого текста и от того его участка, на котором мы в данный момент «находимся». Эта универсальная закономерность естественных текстов называется «закон Хипса»: согласно ему, количество словоформ растет медленнее, чем количество словоупотреблений [4], то есть размер словаря находится в сублинейной зависимости от размера текста.

Скорость уменьшения соотношения словоформ и словоупотреблений (то есть уменьшение вероятности употребления автором нового слова) по мере «развития» текста является специфичным для конкретных жанров и/или авторов параметром [5]. Соответственно, скорость изменения угла наклона графика этого соотношения отражает своеобразие стиля автора произведения. В табл. 2 продемонстрирована динамика изменения TTR в романе Воннегута при последовательном его измерении по мере «нарастания» текста, а на рис. 1 эта динамика представлена графически (все графики и статистические расчеты выполнены в OpenOffice.org Calc).

* Types to tokens ratio.

Таблица 2

Динамика изменения TTR в тексте оригинала

Глава	Словоформы	Словоупотребления	TTR
1	1411	5494	0,26
1+2	2640	12925	0,2
1+2+3	3300	17588	0,19
1+2+3+4	3742	21125	0,18
1+2+3+4+5	4849	32341	0,15
1+2+3+4+5+6	5187	36697	0,14
1+2+3+4+5+6+7	5320	38373	0,14
1+2+3+4+5+6+7+8	5700	43059	0,13
1+2+3+4+5+6+7+8+9	6156	49589	0,12
Весь роман	6247	50848	0,12

Динамика соотношения словоформ и словоупотреблений для романа Курта Воннегута «Slaughterhouse Five...»

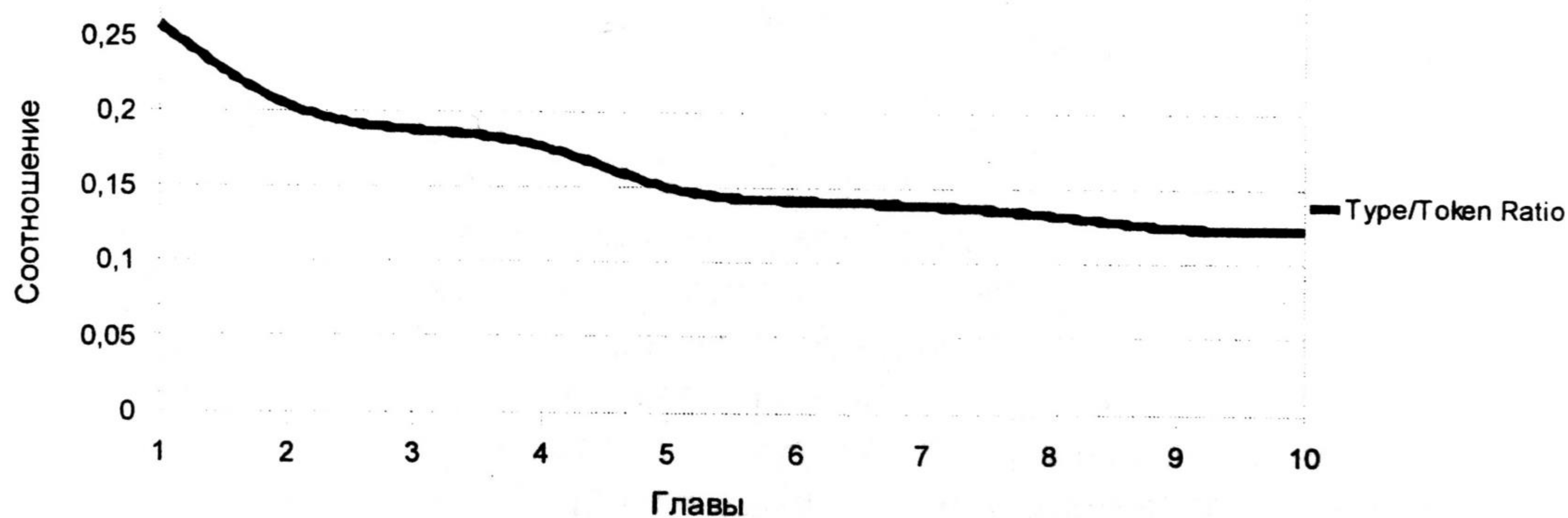


Рис. 1. График динамики TTR в тексте оригинала

Таким образом, мы наблюдаем плавное падение доли уникальных слов (types) с 25% словоупотреблений (tokens) в первой главе до 12% к концу романа, причем наиболее резкое падение наблюдается вплоть до конца 5 главы, то есть до середины романа. Это скорее всего обусловлено тем, что во второй половине текста главы содержат довольно много уникальных слов. Тем не менее этот факт не изменяет общей картины постоянного уменьшения вероятности употребления нового слова.

Рис. 2 наглядно иллюстрирует, насколько рост количества словоформ отстает от роста количества словоупотреблений.

Предположим, что функция, задающая динамику роста числа словоформ в зависимости от числа словоупотреблений, является инвариантом при переводе. Мы можем предварительно проверить это предположение, вычислив соответствующие параметры для русского перевода романа «Slaughterhouse-Five». График роста количества уникальных словоформ для текста оригинала представлен на рис. 3.

В переводе Р. Райт-Ковалевой 12040 словоформ и 41596 словоупотреблений. Таким образом, их соотношение составляет 0,29. Количество словоформ, употребленных лишь один раз (*hapax legomena*) составляет 8188, следовательно, 68% от всех словоформ.

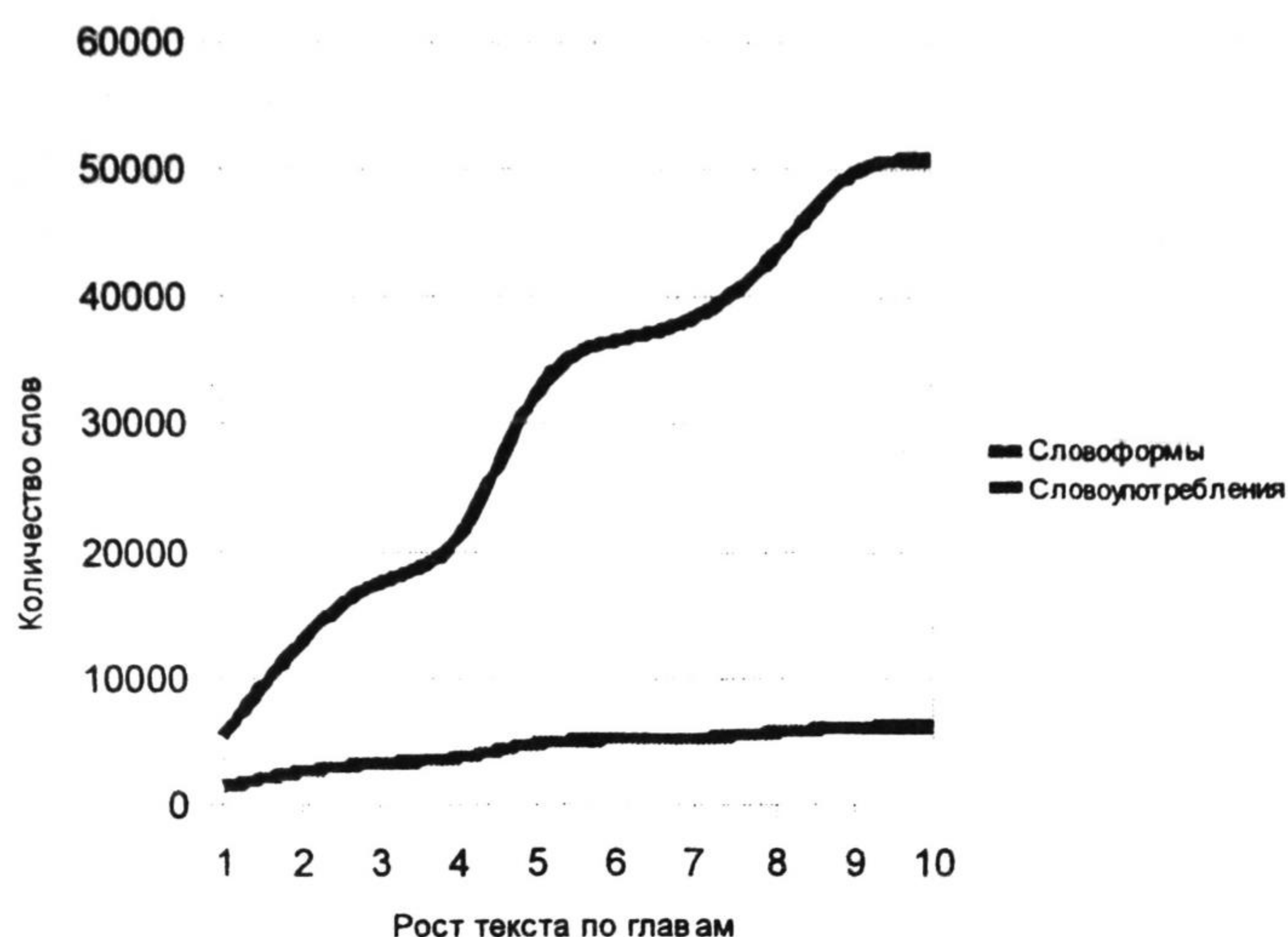


Рис. 2. Рост числа словоформ и словоупотреблений по мере развития текста оригинала

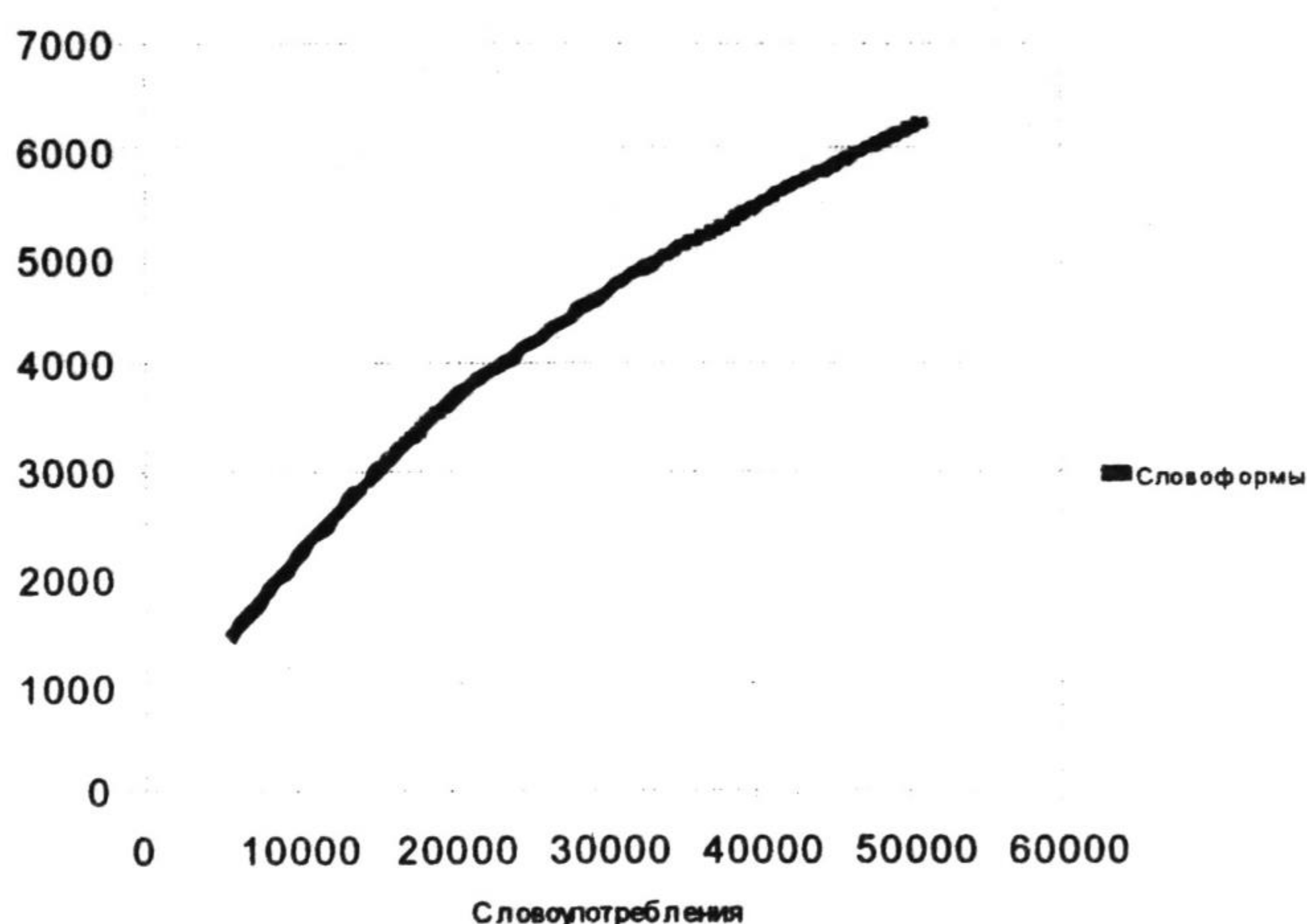


Рис. 3. Рост числа словоформ в тексте оригинала

Отметим, что закон Ципфа продолжает соблюдаться и в русском тексте — частота уменьшается обратно пропорционально рангу словоформы, а наиболее часто употребляемыми словоформами являются служебные слова.

Представляет интерес тот факт, что текст русского перевода на 20% короче, чем текст оригинала (то есть содержит на 20% меньше словоупотреблений). Похоже, что такое соотношение длин характерно для англо-русских переводов вообще, о чем пишет в своей статье М.Н. Михайлов [6]. Это противоречит гипотезе Ю. Найды и Ч. Табера о том, что хороший перевод всегда длиннее оригинала. Между тем исследуемый перевод, определенно, «хороший», что подтверждается его огромной популярностью и словами самого К. Воннегута о профессионализме Р. Райт-Ковалевой (например, в статье *Invite Rita Rait to America!* из книги эссе Воннегута). Мы отсылаем заинтересованных вопросом о длине перевода к вышеупомянутой статье М.Н. Михайлова, а сами вернемся к лексико-частотным характеристикам текста.

В два раза большее по сравнению с оригиналом количество словоформ объясняется синтетическим характером русского языка и следующим из него наличием у большинства слов нескольких морфологических вариантов. Для программы анализа корпуса «Джон» и «Джону» суть разные словоформы. Это по-

влияло и на соотношение TTR, которое составляет 0,29, в отличие от 0,12 у оригинала (или 0,49 против 0,29, если сравнивать нормализованные величины).

На первый взгляд, лексикостатистические зависимости в переводе значительно отличаются от оригинала — число словоформ больше, число словоупотреблений, напротив, меньше, соотношение их, естественно, иное, и так далее. Но не будем спешить и проверим это математически.

Составим табл. 3 для соотношений словоформ и словоупотреблений отдельно в каждой из глав перевода, аналогично таблице для оригинала.

Таблица 3

Лексикостатистические параметры глав перевода

Глава	Словоформы	Словоупотребления	TTR
1	2053	4513	0,45
2	2645	6002	0,44
3	1804	3673	0,49
4	1501	2908	0,52
5	3806	9315	0,41
6	1698	3508	0,48
7	730	1290	0,57
8	1851	3919	0,47
9	2405	5364	0,45
10	617	1004	0,61
Весь роман	12040	41596	0,29 (нормализованное значение 0,49)

Как можно видеть, несмотря на то, что абсолютные значения соотношения TTR у полного текста перевода и отдельных глав больше, чем у оригинала (по абсолютным значениям TTR оригинал лишь в наиболее лексически богатой главе 10 догоняет наименее лексически богатую главу перевода 5), относительные значения TTR-соотношений глав сохраняются. По-прежнему наибольшей долей словоформ характеризуются 7 и 10 главы, а наименьшей — 5 и так далее.

Проверим, выполняется ли для перевода закон Хипса. В табл. 4 представлена динамика изменения TTR в тексте перевода.

Таблица 4

Динамика изменения TTR в тексте перевода

Главы	Словоформы	Словоупотребления	TTR
1	2053	4513	0,45
1+2	4225	10615	0,4
1+2+3	5382	14288	0,38
1+2+3+4	6230	17196	0,36
1+2+3+4+5	8672	26511	0,33
1+2+3+4+5+6	9506	30019	0,32
1+2+3+4+5+6+7	9806	31309	0,31
1+2+3+4+5+6+7+8	10669	35228	0,3
1+2+3+4+5+6+7+8+9	11831	40592	0,29
Весь роман	12040	41596	0,29

Итак, количество словоформ по-прежнему растет медленнее, чем количество словоупотреблений. Посмотрим, сохранились ли объективные параметры этой «медлительности».

На рис. 4 на одной координатной сетке представлены графики изменения соотношения между словоформами и словоупотреблениями в оригинале и в переводе романа.

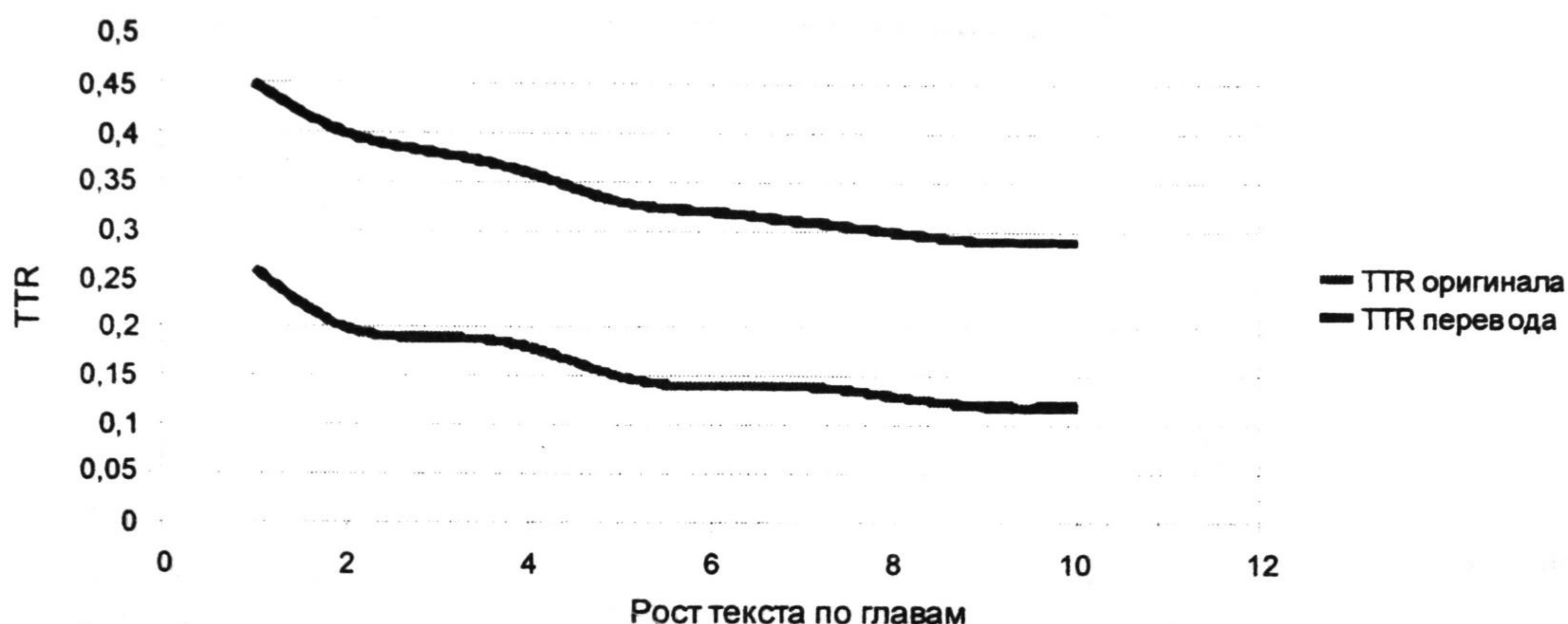


Рис. 4. Соотношение TTR для оригинала и перевода романа

Как можно видеть, график TTR в переводе (черная линия) практически копирует график TTR в оригинале (серая линия), проходя примерно в 0,2 пункта над ним. И действительно, корреляционный анализ показывает 99-процентную вероятность корреляции этих числовых рядов, уже начиная с 3 главы. Это означает, что уже после первых 10000 слов перевода можно говорить о четкой линейной зависимости между TTR отрезков оригинала и соответствующих им отрезков перевода. Следовательно, зная соотношение словоформ и словоупотребления в некоем отрезке текста оригинала, мы можем, даже не читая перевод, однозначно определить, каково будет это соотношение в соответствующем отрезке текста перевода. Это соотношение представляет собой динамический инвариант, переводческую универсалию для данных произведения и переводчика. Переводчик как бы «следует» за оригиналом, сохраняя его статистические параметры, а именно — скорость изменения угла наклона кривой TTR.

Тем не менее, некоторые изменения при переводе, очевидно, произошли, ведь показатели роста числа словоформ не совпадают полностью, что можно видеть на рис. 5. Посмотрим, что же конкретно изменилось в статистических параметрах текста.

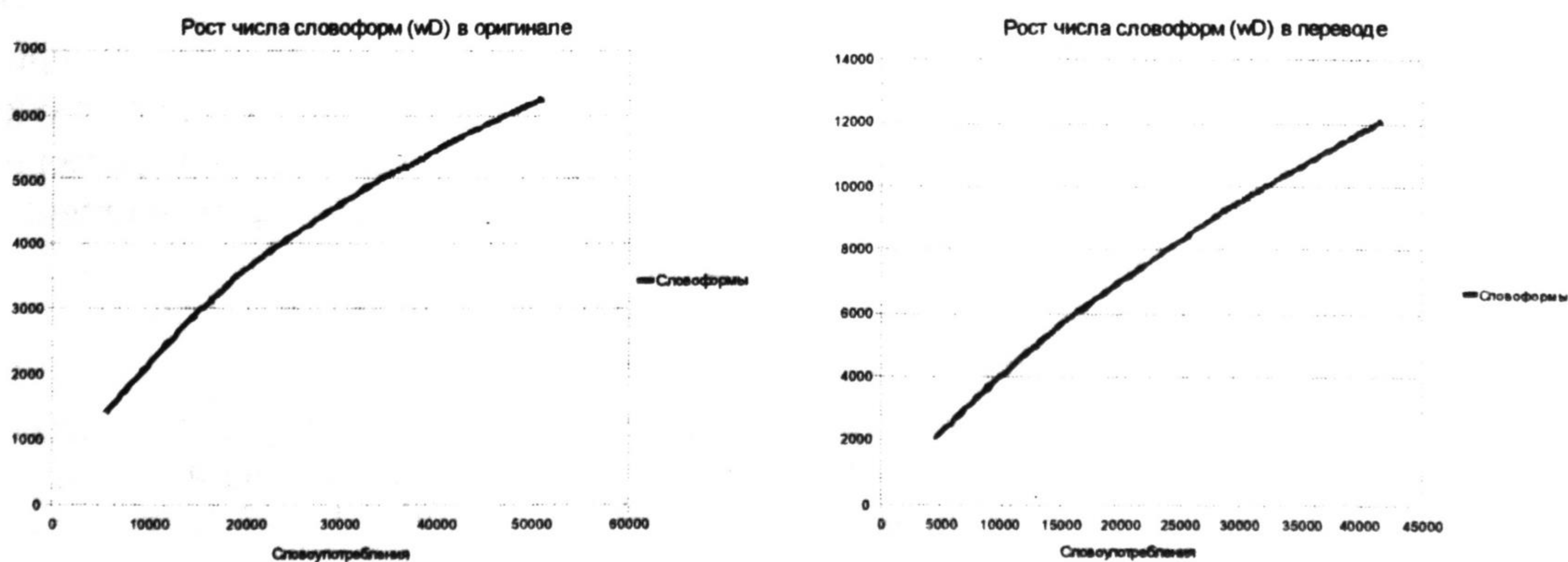


Рис. 5. Рост числа словоформ в оригинале и переводе

Обозначим количество словоформ как wD (Distinct words), а количество словоупотреблений — как $w\gamma$. В целом количество словоформ есть функция от количества словоупотреблений. Согласно закону Хипса, общая вероятность того, что слово встречается в тексте некоторое количество раз, изменяется с развитием текста, а именно — кривая wD становится все более пологой по мере того, как увеличивается $w\gamma$. Изменение этой кривой описывается выражением $Pw\gamma(1)$, то есть вероятность того, что следующее слово будет уникальным (употребленным 1 раз), зависит от $w\gamma$. Зависимость wD от $w\gamma$ — это обычная степенная функция типа $wD = a*(w\gamma)^b$, где a и b — некие коэффициенты, специфичные для данного языка, текста или автора, причем b всегда меньше единицы [4].

Поскольку соотношение числа словоформ и словоупотреблений описывается степенной функцией, мы можем построить кривые степенной регрессии и вычислить значения коэффициентов этой функции в переводе и в оригинале (то есть получить из экспериментальных данных ту функцию, которая вероятнее всего их породила). Степенная регрессия дала нам функцию $wD = 4,98*(w\gamma)^{0,66}$ для текста оригинала и функцию $wD = 2,72*(w\gamma)^{0,79}$ для текста перевода.

Изменился как показатель степени b (с 0,66 до 0,79), так и коэффициент a (с 4,98 на 2,72). Причиной изменения показателя степени, по всей вероятности, является морфологическая система русского языка. Язык этот синтетический, а потому употребление автором новой словоформы, не использовавшейся ранее, более вероятно, чем в аналитическом английском языке. Поэтому показатель степени увеличивается, а следовательно — увеличивается и количество словоформ по отношению к количеству словоупотреблений. Таким образом, скорее всего изменение показателя степени системно для пары «английский-русский языки».

Причина уменьшения коэффициента a пока неясна. Это может быть как воздействие так называемого «голоса переводчика» (его индивидуальных особенностей), так и случайной флуктуацией. Здесь нужны дальнейшие исследования на более широком материале, например, на других переводах Р. Райт-Ковалевой.

Итак, выявлено, что существуют объективные статистические различия между текстами оригинала и перевода, которые можно исследовать и выявлять ответственные за них явления в системе языка перевода или же в самом процессе переводческой деятельности. Но несмотря на эти различия, можно предварительно говорить о том, что на уровне общей динамики соотношения количества словоформ и словоупотреблений разница грамматических систем двух языков нивелируется и переводчик сохраняет некий «динамический лексический инвариант» текста. Визуальным выражением этого инварианта на графике является скорость изменения угла наклона кривой TTR по отношению к горизонтальной оси координат. Инвариантность этого параметра позволяет статистически сравнивать между собой разные переводы (или работы разных переводчиков), а кроме того, дает еще одно направление для дальнейшего изучения перевода как вида человеческой деятельности.

СПИСОК ЛИТЕРАТУРЫ

1. Olohan, M. *Introducing Corpora in Translation Studies*. NY: Routledge, 2004. 220 с.
2. Baker, M. *Corpus Linguistics and Translation Studies: Implications and Applications*. Amsterdam: John Benjamins Publishing Company, 1993. С. 233-250.
3. Zipf, G. *Selective studies and the principle of relative frequency in language*. Cambridge, Massachusetts: Harvard University Press, 1932.

4. Serrano, M.A., Flammini, A., Menczer, F. Beyond Zipf's Law: Modelling the structure of human language. Arxiv.org. 2009. URL: <http://arxiv.org/abs/0902.0606> (от 22.10.2009).

5. Bernhardsson, S., Rocha, L.E.C. da, Minnhagen, P. Size dependent word frequencies and translational invariance of books. Arxiv.org. 2009. URL: <http://arxiv.org/abs/0906.0716> (от 22.10.2009).

6. Михайлов М.Н. Чем длиннее, тем лучше? Как сравнить длины исходного текста и перевода? // Математическая морфология. 2003. Т. 5. Вып. 1. Смоленск: СГМА.

*Марина Ивановна МЕХЕДА —
старший преподаватель
кафедры английского языка
Тюменского государственного университета
mmekheda@mail.ru*

УДК 81-115(111+161.1)

ТЕМАТИЧЕСКАЯ ДИФФЕРЕНЦИАЦИЯ РЕДУПЛИКАТИВОВ В РУССКОМ И АНГЛИЙСКОМ ЯЗЫКАХ

THEMATIC DIFFERENTIATION OF RUSSIAN AND ENGLISH REDUPLICATIVES

АННОТАЦИЯ. В статье рассматриваются различные аспекты тематической дифференциации редупликативных образований в русской и английской субстандартной лексической подсистеме.

SUMMARY. The article deals with thematic structure and pragmatics of Reduplication in the Russian and English languages.

КЛЮЧЕВЫЕ СЛОВА. Редупликация, просторечье, понятийная сфера, информативность.

KEY WORDS. Reduplication, substandard layer of the vocabulary, thematic field, informative value of lexical unit.

Настоящая статья посвящена анализу некоторых аспектов семантической дифференциации редупликативных образований субстандартной лексической подсистемы русского и английского языков. Актуальность данной работы обусловлена отсутствием детальных исследований в области редупликации, как стилистически маркированного способа словообразования, с позиции информативности лексических единиц и языковой экономии, которые определяют создание и функционирование редупликативов в речи. В статье также рассматриваются прагматический и социальный аспекты семантической дифференциации редупликативных образований в русской и английской субстандартной лексике.

Под субстандартной подсистемой мы вслед за В.А. Хомяковым понимаем сложную социально-стилистическую категорию, состоящую из социально-профессионального и экспрессивного просторечия [1]. Известно, что просторечие как пласт лексики, характерное для любого естественного языка, отличается большей свободой словотворчества в связи с отсутствием сдерживающих факторов в виде строгого соответствия литературным нормам словопроизводства, синтаксически нормированной организации высказывания, этических соображений. Именно благодаря разнообразию единиц лексики просторечное высказывание более информативно с точки зрения эмоционально-экспрессивного заряда. Избыточность, синонимическая вариативность и экспрессивность про-