

ŠIAULIŲ UNIVERSITETAS
HUMANITARINIS FAKULTETAS
RUSŲ KALBOS KATEDRA

Živilė Armalytė

INTERNETO TEKSTŲ LIETUVIŲ IR RUSŲ KALBOMIS LEKSIKOS GRE TINAMASIS
LINGVOSTATISTINIS

Magistro darbas

Šiauliai,
2005

**ШЯУЛЯЙСКИЙ УНИВЕРСИТЕТ
ГУМАНИТАРНЫЙ ФАКУЛЬТЕТ
КАФЕДРА РУССКОГО ЯЗЫКА**

Живиле Армалите

Сопоставительное исследование лексики интернетных текстов на
русском и литовском языках методами лингвистической
статистики

Магистерская работа

**Шяуляй,
2004**

Содержание

1. Введение. Обоснование выбора темы.....	7
2. Цели и задачи исследования.....	8
3. Обзор литературы по проблеме исследования.....	9
4. Характеристика подготовительных работ.....	12
5. Лексикостатистический анализ разных текстов на русском языке.....	13
5.1. Лингвостатистическая характеристика тематических текстов.....	13
5.1.1. Тексты Ветхого Завета.....	13
5.1.2. Тексты по художественной литературе.....	15
5.1.3. Тексты по медицине.....	16
5.1.4. Тексты по механике и автомобилизму.....	17
5.1.5. Тексты о войне.....	18
5.1.6. Тексты о женщинах.....	19
5.1.7. Тексты о животном мире.....	20
5.1.8. Тексты о политике.....	21
5.1.9. Тексты о банковском деле.....	21
5.1.10. Тексты о спорте.....	22
5.1.11. Параметры и лексический состав сводного словника.....	23
5.1.12. Анализ высокочастотной заимствованной лексики.....	24
6. Лексикостатистический анализ разных текстов на литовском языке.....	35
6.1. Лингвостатистическая характеристика тематических текстов.....	35
6.1.1. Тексты Ветхого Завета (на литовском языке).....	36
6.1.2. Тексты художественной литературы (на литовском языке).....	37
6.1.3. Тексты о медицине (на литовском языке).....	38
6.1.4. Тексты по автомобилизму (на литовском языке).....	39
6.1.5. Тексты о войне (на литовском языке).....	41
6.1.6. Тексты о женщинах (на литовском языке).....	42
6.1.7. Тексты о сейме (на литовском языке).....	43
6.1.8. Тексты о политике (на литовском языке).....	44
6.1.9. Тексты о банковском деле (на литовском языке).....	46

6.1.10. Тексты о спорте (на литовском языке).....	47
6.1.11. Параметры и лексический состав сводного словаря.....	48
6.1.12. Анализ высокочастотной лексики в текстах на литовском языке.....	50
7.Сопоставительный анализ высокочастотной русско-литовской лексики.....	56
8. Сопоставительный анализ низкочастотной русско-литовской лексики.....	57
9. Обобщение результатов исследования. Выводы.....	57
10.Библиография.....	61
11.Резюме.....	62
12.Таблицы и графики.....	63

План работы:

1. Введение. Обоснование выбора темы. Актуальность проблемы исследования.
2. Цели и задачи исследования.
3. Обзор литературы по проблеме исследования.
4. Характеристика подготовительных работ.
5. Лексикостатистический анализ разных текстов на русском языке:
 - 5.1. Лингвостатистическая характеристика тематических текстов:
 - 5.1.1. Тексты Библии (Ветхого Завета);
 - 5.1.2. Тексты по художественной литературе;
 - 5.1.3. Тексты по медицине;
 - 5.1.4. Тексты по автомобилизму;
 - 5.1.5. Тексты о войне;
 - 5.1.6. Тексты о женщинах;
 - 5.1.7. Тексты о животном мире;
 - 5.1.8. Тексты о политике;
 - 5.1.9. Тексты о банковском деле;
 - 5.1.10. Тексты о спорте;
 - 5.1.11. Параметры и лексический состав сводного словаря;
 - 5.1.12. Анализ высокочастотной заимствованной лексики.
6. Лексикостатистический анализ разных текстов на литовском языке:
 - 6.1. Лингвостатистическая характеристика тематических текстов:
 - 6.1.1. Тексты Библии (Ветхого Завета);
 - 6.1.2. Тексты по художественной литературе;
 - 6.1.3. Тексты по медицине;
 - 6.1.4. Тексты по автомобилизму;
 - 6.1.5. Тексты о войне;
 - 6.1.6. Тексты о женщинах;
 - 6.1.7. Тексты о сейме;
 - 6.1.8. Тексты о политике;
 - 6.1.9. Тексты о банковском деле;
 - 6.1.10. Тексты о спорте;

6.1.11. Параметры и лексический состав сводного словаря.

6.1.12. Анализ высокочастотной заимствованной лексики.

7. Сопоставительный анализ высокочастотной русско-литовской лексики.

8. Сопоставительный анализ низкочастотной русско-литовской лексики.

9. Обобщение результатов исследования. Выводы.

10. Список литературы.

11. Перечень источников.

12. Приложения:

12.1. Фрагменты частотных словников русских тематических текстов.

12.2. Фрагменты частотных словников литовских тематических текстов.

12.3. Таблицы и графики.

1. Введение. Обоснование выбора темы. Актуальность проблемы исследования.

Совершенствование и распространение персональных компьютеров даёт возможность внедрять эту технику в процесс научных исследований, в частности, в гуманитарные области. В силу этого компьютеры используются в лексикографии, лингвостатистике, лингводидактике и других частных областях языковедения и методике преподавания языков.

С помощью компьютера можно выполнять исследования при наличии чрезвычайно большого материала. Таким образом, появляется возможность получения статистически достоверных результатов на базе очень большой выборки, что, например, актуально при составлении частотных словарей по текстам разной тематики.

Известно, что частотные словари до последних десятилетий предыдущего столетия прошлого тысячелетия составлялись на базе одного миллиона словоупотреблений. Такая выборка считалась статистически достоверной.

В источниках научной литературы указывается, что авторы стремились к отражению характеристик всей лексической системы языка. Поэтому они старались охватывать разные функционально речевые сферы; допустим, в «Частотном словаре русского языка» (ЧСРЯ) охвачено: художественная проза – 25,4%; драматургия – 27,2%; научные и публицистические тексты – 23,6%; газетные и журнальные тексты – 23,8%.

Однако лексическая система любого языка – это явление, сильно изменяющееся во времени (вследствие динамических изменений жизни общества).

Поэтому встаёт вопрос о том, как часто надо составлять частотные словари, чтобы иметь сведения, важные для теории и практики: ведь нужно знать, какие слова можно отнести к ядру лексической системы, а какую – к периферии. Помимо этого, в научной литературе не описан фактологический материал и результаты, зависящие от объёма выборки и тематического ранжирования текстов. Известно лишь, что наиболее употребительные союзы, предлоги, некоторые местоимения, наречия и т.д. являются высокочастотными элементами любой тематической выборки текстов. Но почти нет сведений о том, какими должны быть минимальные и максимальные параметры

исследуемых текстов, составляющих статистически достоверный словарь как зеркальное отражение объёма и состава лексической системы любого языка. Иначе говоря, при составлении частотных словарей всегда речь идёт о двух самых важных факторах:

а) о величине корпуса текстов, или сумме выборок;

б) о тематике текстов и их роли при образовании универсального частотного словаря.

Следовательно, актуальность обсуждаемой темы становится очевидной, особенно в условиях, когда для исследования формируется десять интернетных текстов (корпусов), объём каждого из которых – сто тысяч словоупотреблений (тексты Ветхого Завета и художественной литературы; научные тексты и тексты по автомобилизму; тексты о войне, женщинах, животном мире; политические и экономические тексты, а также тексты о спорте).

Для целей сопоставления такие же виды работ предусмотрены и по текстам родного (литовского) языка.

Ожидаемые результаты исследования могут характеризоваться элементами новизны, так как аналогичные результаты детально не описываются в доступных нам источниках специальной лингвистической литературы.

2. Цели и задачи исследования.

Основная цель – это определение статистических параметров десяти текстов разной тематики, но одинаковых по объёму (по 100000 словоупотреблений в каждом из них). Так как исследование должно было быть сопоставительным, то в таких же пределах предполагалось выполнить аналогичную работу по текстам на родном (литовском) языке. Полученные результаты должны были сверяться на базе русского (рабочего) языка данного исследования.

Однако были и дополнительные цели:

а) составить словарь по десяти темам, объём текстов которых, как указывалось, составляет 100000 словоупотреблений (предположительная суммарная величина выборки – 1 000 000 млн. словоупотреблений);

б) определить роль тематики текстов при наличии одинаковой их выборки (100 000 словоупотреблений в каждой).

Для достижения намеченных целей надо было решить следующие задачи:

- 1) ознакомиться с литературой по проблеме исследования;
- 2) заготовить десять файлов (текстов разной тематики на русском языке) и дополнительно десять файлов (текстов соответствующей тематики на литовском языке), чтобы общая выборка составила 1000000 словоупотреблений на русском языке и столько же – на литовском;
- 3) составить словники на базе вышеупомянутого объёма лексического материала, позаимствованного из Интернета;
- 4) произвести простое сложение (суммирование) десяти текстов разной тематики для получения сводного словаря;
- 5) произвести сопоставительный анализ высокочастотной лексики текстов аналогической тематики русского и литовского (в переводе на русский) языков;
- 6) проанализировать высокочастотные лексические заимствования;
- 7) обобщить результаты сопоставительного исследования в целом, сделать выводы;
- 8) оформить приложения и лексикостатистическое исследование в целом.

3. Обзор литературы по проблеме исследования

Известно, что всякое исследование начинается с познания того, что было сделано другими учёными до момента определения собственной темы научного труда. Ведь всякая проблема и её исследование имеет связь с прошлым, с историей целого ряда вопросов, связанных между собой.

Б.Н.Головин в книге “Язык и статистика” (1971) указывает, что древние учёные Александрийской поры, желая более эффективно обучать молодых людей по текстам священных книг, производили подсчёты словоупотреблений, чтобы таким образом дифференцировать процесс обучения, т.е. отбора наиболее употребительных и самых важных слов, составляющих ядро лексического наполнения изучаемых текстов.

Конечно, эта работа производилась вручную. С тех пор прошло много времени, пока учёные и инженеры-конструкторы приблизительно к середине прошлого столетия создали электронно-вычислительные машины, которые позволяли выполнять подсчёты на материале очень больших баз данных.

Так, в 1948 г., компанией IBM была подарена одна из первых электронно-вычислительных машин для лингвистической лаборатории итальянского университета с заданием произвести лексикостатистический анализ текстов Библии. Условно можно считать, что в середине прошлого века зародилась новая отрасль лингвистики, базирующаяся на методах математической статистики. Об этом можно найти сведения в книге П.Н.Алексеева под названием “Статистическая лексикография” (1975).

Помимо этого, в “Лингвистическом энциклопедическом словаре” (1990) указывается, что “математическая лингвистика – это дисциплина, предметом которой является разработка формального аппарата для писания, строения естественных и некоторых искусственных языков, возникшая в 50-ых годах двадцатого века. Одним из главных стимулов появления математической лингвистики послужила назревшая в языкознании потребность уточнения его основных понятий...”. «Математическая лингвистика развивается в тесном взаимодействии с языкознанием» (207). Помимо этого, в данном источнике определяется понятие *лингвостатистика* (269). Указывается, что в широком смысле – это область применения статистических методов в языкознании. Также в указанном энциклопедическом словаре (231) определяются *количественные методы* в языкознании. Это использование подсчётов и измерений при изучении языка и речи. Отметим также, что в этом смысле в русском языкознании известны труды Н.Андреева (“Статистико-комбинаторные методы в теоретическом и прикладном языковедении”(1967), А.Шайкевича (“Дистрибутивно-статистический анализ в семантике”), Р.Пиотровского (совместно с другими авторами (“Математическая лингвистика”(1977). Кроме того, по переводам с французского и английского языков на русский язык нам известны труды зарубежных учёных по статистике, например: P.Guiraud “Problemes et methodes de la statistique” (1959).

Следует отметить также книги, которые имеют связь с данной проблематикой: “Статистика речи и автоматический анализ текста” (1974); Р.Пиотровский “Текст. Машина. Человек.” (1975); “Инженерная лингвистика и

теория языка” (1979); “Вычислительная лингвистика” (1975); “Статистика и функциональные стили языка” (1977). Указанные книги в большей степени характеризуют общие и частные проблемы по лингвистической статистике.

Однако к нашему исследованию относятся труды и по компьютерной лингвистике. В этом плане можно указать такие книги, которые имеют прямое отношение для целей сопоставления и сверок в нашей научной работе, например: “Частотный словарь русского языка” (ЧСРЯ) под ред. Л.Засориной; “Частотный словарь современного русского языка” под ред. Л.Лённгрена (1993); L.Grumadienė, V.Žilinskienė “Dažninis dabartinės rašomosios lietuvių kalbos žodynas” (1998); “Latviešu valodas biežuma vardnīca” (“Частотный словарь латышского языка”) (1973); “Частотный словарь польского языка” (“Słownik frekwencyjny polszczyzny współczesnej”, 1-2 t.) (1990).

По указанным источникам можно определить методологию подготовки такого рода словарей, что полезно и для нашего научного исследования. Отметим также научные труды Ю.Корсакаса, например: “Лексика дефиниций” (1998), “Частотная лексика” (1999), “Текст и частота слова” (2004).

В этих трудах освещаются вопросы анализа лексического наполнения разных текстов и дифференциация лексики по частоте употребления слов, что соотносится с прикладными аспектами – активным и пассивным словарём, подлежащим или не подлежащим к включению в процесс обучения. Отметим также, что аналогичные проблемы освещаются в книге В.Кузиной под названием “3000 наиболее употребительных слов латышского языка в соотношении с аналогами в русском, немецком и английском языках” (1998).

На родном (литовском) языке важным источником знаний является 24-ый том научных трудов под названием “Darbai ir dienos” учёных университета Витаутаса Великого города Каунас. В указанном сборнике научных трудов имеется целый ряд статей, в числе которых прямое отношение к нашему исследованию имеет статья Р.Марцинкявичене о принципах создания корпуса текстов для решения разных лингвистических задач, в частности, для определения частот словоупотреблений. В центре компьютерной лингвистики указанного университета создан чрезвычайно большой корпус текстов на литовском языке, поэтому с показателями величины индексов можно сверять (и

проверять) частоты по любому слову литовского языка. Следовательно, описанный корпус имеет прямое отношение к нашей работе.

Знакомство и изучение научной литературы по проблеме исследования помогло в построении собственного плана работы и реализации многих аспектов нашего научного труда.

4. Характеристика подготовительных работ

Чтобы выполнить определённую научную работу, надо выполнить целый ряд подготовительных работ. В нашем случае надо было:

во-первых, ознакомиться с немалым объёмом литературы по избранной проблеме;

во-вторых, выполнить большой объём работ, связанных с подготовкой текстов для сопоставительного лексико-статистического анализа;

в-третьих, составить словники, лингвостатистические таблицы и графики по текстам разной тематики на русском и литовском языках;

в-четвёртых, выполнить интервальный анализ лексического состава тематических словников;

в-пятых, перевести 500 наиболее частотных знаменательных слов с родного языка на русский, на базе которого производился сопоставительный анализ лексического состава словников текстов разной тематики;

в-шестых, оформить результаты исследования в целом.

Подготовительные виды работ имели разнообразный характер. Изучение литературных источников, можно сказать, отличается традиционным отбором статей и книг разных авторов на литовском, русском и английском языках. Трудность заключалась в том, что многие источники значатся в библиографических списках, однако отсутствуют в библиотеках страны и с трудом доступны по межбиблиотечным обменам.

В начале исследования трудности возникли и в процессе формирования текстов разной тематики, так как многие сайты Российской Федерации в Интернете сейчас закрыты или с большим трудом доступны для иностранных граждан. Помимо этого, некоторые темы характеризуются весьма малым числом

текстов, поэтому аналогичную тематику приходилось дополнять материалами из других сайтов разных стран.

Много времени ушло и на лексико-статистическую обработку текстов, составление словарей, статистических таблиц и графиков.

Однако в целом исследование вызывает интерес, так как его результаты являются конкретными и отличается некоторой новизной (они в печати нашей страны ещё не публиковались).

5. Лексикостатистический анализ разных текстов на русском языке

Как уже указывалось, для сопоставительного исследования лексики была определена условная (теоретическая) совокупность текстов, объём которой – не менее 1000000 словоупотреблений: ведь надо было получить относительную достоверность статистических данных и лексический состав частотного словаря.

Для того чтобы получить относительно *универсальную* совокупность текстов, надо иметь, как мы определили для данного исследования, десять одинаковых или схожих по параметрам подтекстов разной тематики.

Далее рассмотрим параметрические данные и состав словарей по каждой исследуемой теме.

5.1. Лингвостатистическая характеристика тематических текстов

В данном разделе будут описаны лексико-статистические результаты по отдельным текстам на русском языке.

Рассмотрим статистические данные по каждой теме текстов в отдельности.

5.1.1. Тексты Ветхого Завета

Как уже указывалось, данная тематическая совокупность текстов позаимствована из Интернета. Ознакомимся со статистическими их характеристиками, чтобы на таком основании можно было выполнять сопоставления.

Первоначальные статистические параметры данных текстов были получены по таблице “Word Count”. Они были следующими:

- а) объём текстов – 205 страниц;
- б) всего слов – 113 461;
- в) они составлены из 524058 знаков без интервалов.

Следовательно, средняя длина слова состоит из 4,62 знаков. Иначе говоря, средняя длина слова текстов Библии (Ветхого Завета) – 4-5 букв.

$$(524\ 058 : 113\ 461)$$

После лексикостатистического анализа данных текстов (с учётом частичного редактирования словника – удаления из текстов чисел и “непонятных” обозначений), на основе компьютерной программы “Лингвостатистика” был составлен частотный перечень слов (по убывающим частотам), а также таблицы, отражающие статистическую структуру словника.

В итоговом варианте указывается, что в текстах Библии (Ветхом Завете) насчитывается 107334 словоупотребления; длина, или объём, словника – 12588 разных слов и словоформ. Следовательно, индекс итерации (среднее число употребительности слов) составляет – 8,53 (107334 : 12588).

Коэффициент лексического разнообразия – 0,1173 (12588 : 107334).

В числе наиболее употребительных слов были: *Господь-1263; Бог-813; земля-821; сын-549; день-328; народ-179; год-177; имя-163; отец-154; священник-146, лицо-81; скот-81; сердце-80; дочь-77; жена-77; одежда-77; брат-73; вода-73; слово-72, кровь-121 и т.д.*

Наиболее частотными были следующие имена собственные: *Иаков-92; Авраам-45; Исаак-45; Египет-76; Исав-42; Аарон-40; Израиль-38; Иов-37; Иосиф-37; Лаван-35; Моисей-35 и т.д.*

Статистическая структура представлена в частотной таблице (см. №1).

Таблица №1

ЧАСТОТНАЯ СТРУКТУРА ТЕКСТОВ «ВЕТХОГО ЗАВЕТА»

Ран-ги	Частотная группа	Частота в словнике	Процент	Количество словоупотр.	Процент
1	1 – 4	10093	80,1795	16134	15,0316
2	5 – 9	1219	9,6838	7842	7,3062
3	10 – 100	1145	9,0960	30729	28,6293

4	101 – 1000	125	0,9930	34398	32,0476
5	>1000	6	0,0477	18231	16,9853

Следовательно, наибольшую долю словника составляют низкочастотные словоупотребления – свыше 80%.

5.1.2. Тексты художественной литературы

В текстах русской художественной литературы насчитывается по подсчетам программы “Microsoft Word” (Word Count) 104741 словоупотребление. Эта совокупность составлена из 588583 знаков. Таким образом, средняя длина слова – 5,62 знака. В предварительно отредактированном варианте текстов с помощью компьютерной программы «Лингвостатистика» осталось 102249 словоупотреблений. Объём словника – 25004 разных слова и словоформы. Следовательно, индекс итерации (среднее число употребительности слов) составляет – 4,06 (102249 : 25004). Коэффициент лексического разнообразия – 0,2484 (25004 : 102249).

В числе наиболее употребительных слов были: *человек-131; рука-105; дом-102; смерть-90, Мать-77; год-71; жизнь-119; день-61; рот-61; лицо-59; глаз-111; вещь-108; кровать-54; окно-54; дом-49; голова-47; друг-47; тело-47; дверь-46; ветер-43 и т.д.*

Наиболее частотными были следующие имена собственные: *Иван-377; Мухин-355; Васильевич-138; Ивановна-116; Антонина-112, Катюша-58; Мерзляков-52; Том-50; Васильевич-48, Коляда-20; Сокуров-20; Борис-19; Детин-19; Иван-37; Калитников-18; Пальмирова-18; Прасковья-18 и т.д.*

Данные представлены в виде результатов частот по их группам (см. табл. №2).

Таблица №2

ЧАСТОТНАЯ СТРУКТУРА ТЕКСТОВ ПО РУССКОЙ ЛИТЕРАТУРЕ

Ранги	Частотная группа	Частота в словнике	Процент	Количество словоуп.	Процент
1	1 – 4	22208	88,8178	33058	32,3309
2	5 – 9	1566	6,2630	9937	9,7184
3	10 – 100	1136	4,5433	24869	24,3220
4	101 – 1000	87	0,3479	20890	20,4305
5	>1000	7	0,0280	13495	13,1982

Как видно, и в данном тематическом словнике наибольшую долю составляют низкочастотные словоупотребления – почти 89%.

5.1.3. Лексикостатистическая характеристика текстов по медицине

После лексикостатистического анализа данных текстов на основе компьютерной программы “Лингвостатистика” был составлен частотный перечень слов (по убывающим частотам), а также таблицы, отражающие статистическую структуру словника.

В итоговом варианте указывается, что в текстах медицины насчитывается 97186 словоупотреблений; длина, или объём, словника – 15432 разных слов и словоформ. Следовательно, индекс итерации (среднее число употребительности слов) составляет – 6,3 (97186 : 15432). Коэффициент лексического разнообразия – 0,1587 (15432 : 97186).

В числе наиболее употребительных слов были: *матка-369; женщина-449; лечение-468; эндометрия-267; беременность-239; течение-220; время-215; шейка-198; заболевание-275; яичник-172; цикл-164; кровь-161; день-(1)150; функция-147; изменение-133; нарушение-131; тело-128; развитие-125; год-121; терапия-121 и т.д.*

Статистические данные представлены в виде частотных групп (см. табл. №3).

Таблица №3

ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ ПО МЕДИЦИНЕ

Ранги	Частотная группа	Частота в словнике	Процент	Количество словоуп.	Процент
1	1 – 4	12051	78,0910	21728	22,3571
2	5 – 9	1764	11,4308	11437	11,7682
3	10 – 100	1532	9,9274	36910	37,9787
4	101 – 1000	81	0,5249	17603	18,1127
5	>1000	4	0,0259	9508	9,7833

В медицинских текстах повторяемость терминов весьма заметна, поэтому низкочастотная группа здесь несколько меньше и составляет чуть более 78%.

5.1.4.Тексты по механике и автомобилизму

После лексикостатистического анализа текстов с помощью программы “Лингвостатистика” был составлен частотный перечень слов (по убывающим частотам), а также таблицы, отражающие статистическую структуру словника.

В итоговом варианте в текстах по механике и автомобилизму насчитывается 122524 словоупотребления; длина, или объём, словника – 8587 разных слов и словоформ. Следовательно, индекс итерации (среднее число употребительности слов) составляет – 14,3 (122524 : 8587). Коэффициент лексического разнообразия – 0,0709 (8587 : 122524).

В числе наиболее употребительных слов были: *страхование-1760; средство-2532; движение-1566; автомобиль-1796; безопасность-896; закон-1354; ответственность-814; соответствие-814; статья-758; объединение-754; выплата-676; случай-630; машина-1016; вред-588; власть-562; страховщик-876; право-474; ответственность-462; год-458; счёт-428 и т.д.*

Наиболее частотными были следующие имена собственные: *Москва-318; Toyota-304; Jetta-224; Россия-224, ГАИ-176, Lanos-140, Япония-94, Владивосток-70; Corolla-68; Nissan-60 и т.д.*

Данные представлены в виде результатов по частотным группам (см. табл. №4).

Таблица №4

ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ ПО АВТОМОБИЛИЗМУ И ТЕХНИКЕ

Ранги	Частотная группа	Частота в словнике	Процент	Количество словоуп.	Процент
1	1 – 4	644	7,2777	1592	0,4478
2	5 – 9	972	10,9843	7536	2,1195
3	10 – 100	6730	76,0538	154686	43,5064
4	101 – 1000	473	5,3452	112814	31,7296
5	>1000	30	0,3390	78920	22,1967

Здесь получена частотная структура словника чрезвычайно своеобразная: низкочастотные словоупотребления составляют самую малую долю словника – только чуть более 7%, а высокочастотные – самую большую (почти 82%). Такое

явление было получено за счет того, что в текстах чрезвычайно много терминов, которые повторяются множество раз.

5.1.5.Тексты о войне

После лексикостатистического анализа текстов на основании компьютерной программы “Лингвостатистика” был составлен перечень слов (по убывающим частотам), а также таблицы, отражающие статистическую структуру словника.

В итоговом варианте, в текстах о войне насчитывается 100000 словоупотреблений; длина, или объём, словника – 23212 разных слов и словоформ. Следовательно, индекс итерации (среднее число употребительности слов) составляет – 4,3 (100000 : 23212). Коэффициент лексического разнообразия – 0,2321 (23212 : 100000).

В числе наиболее употребительных знаменательных слов были: *общество-316; армия-294; время-196; война-162; войско-205; власть-98; террор-98; страна-94; человек-90; год-83; правительство-82, организация-59; случай-59; дело-53; день-51; вопрос-48; население-47; анархия-45; безопасность-45; конфликт-44 и т.д.*

Наиболее частотными были следующие имена собственные: *Россия-116; США-105, Германия-57; МК-57; Чечня-57, Грозный-37; Израиль-32; Германия-31; Дудаев-29; Франция-22; НАТО-21; Израиль-20; Москва-19; Останкино-19 и т.д.*

Данные представлены в виде статистических результатов по частотным группам (см. табл. №5).

Таблица №5

ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ О ВОЙНЕ

Ранги	Частотная группа	Частота в словнике	Процент	Количество словоуп.	Процент
1	1 – 4	20172	86,9033	30083	30,0830
2	5 – 9	1679	7,2333	10799	10,7990
3	10 – 100	1281	5,5187	28569	28,5690
4	101 – 1000	74	0,3188	17848	17,8480
5	>1000	6	0,0258	12701	12,7010

Весьма четко просматривается тематическая направленность словника. Однако и здесь наибольшую долю составляют низкочастотные смловоупотребления – около 87%, а высокочастотные – лишь 5,8%.

5.1.6.Тексты о женщинах

После лексикостатистического анализа текстов с помощью программы “Лингвостатистика” был составлен частотный перечень слов, а также таблицы, отражающие статистическую структуру словника.

В текстах о женщинах насчитывается 103182 словоупотребления; длина словника – 26341 разное слово и словоформа. Следовательно, индекс итерации (среднее число употребительности слов) составляет – 3,9 (103182 : 26341). Коэффициент лексического разнообразия – 0,2552 (26341 : 103182).

В числе наиболее употребительных слов были: *женщина-932; мужчина-301; песок-198; время-196; жизнь-122; человек-200; жизнь-78; год-77; история-74; мужчина-73; глаз-72; время-60; движение-59; день-59; организация-56; дело-55; тело-54; материнство-53; правда-53; вода-50 и т.д.*

Наиболее частотными были следующие имена собственные: *Россия-37; Карелия-36; США-15; Москва-12 и т.д.*

Результаты анализа представлены в частотной таблице (см. №6).

Таблица №6

ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ О ЖЕНЩИНАХ

№	Частотная группа	Частота в словнике	Процент	Количество словоупотр.	Процент
1	1 – 4	23541	89,3702	33818	32,7751
2	5 – 9	1562	5,9299	10009	9,7003
3	10 – 100	1156	4,3886	26986	26,1538
4	101 – 1000	76	0,2885	19920	19,3057
5	>1000	6	0,0228	12449	12,0651

По наиболее частотным словоупотреблениям видно, что здесь действительно представлен словник тематических текстов о женщинах. Но и в данном случае «раговоры велись» на всевозможные темы с использованием разнообразнейших словоупотреблений, что отразилось на частотной структуре словника: наибольшую долю словника составляет низкочастотная лексика – свыше 89%; лишь 4,5% составляют высокочастотные словоупотребления.

5.1.7.Тексты о животном мире

После лексикостатистического анализа текстов с помощью программы “Лингвостатистика” был в очередной раз составлен частотный перечень слов, а также таблицы, отражающие статистическую структуру словника.

В итоговом варианте в текстах о животном мире насчитывается 128279 словоупотреблений; длина словника –25343 разных слова и словоформы.

Следовательно, индекс итерации (среднее число употребительности слов) составляет – 5,1 (128279 : 25343).

Коэффициент лексического разнообразия – 0,1975 (25343 : 128279).

В числе наиболее употребительных слов были: *время-251; тело-217; вид-197; хвост-165; жизнь-159; самка-146; птица-244; животное-131; змея-129; день-116; год-111; земля-109; длина-106; самец-106; зверь-80; акула-79; часть-72; период-71; вид-69; голова-69 и т.д.*

Наиболее частотными были следующие имена собственные: *Азия-143; Африка-81; Россия-79; Индия-77; Америка-117; Европа-54; Австралия-28; Китай-26 и др.*

Данные о статистической структуре представлены в частотной таблице (см.№7).

Таблица №7

ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА О ЖИВОТНОМ МИРЕ

№	Частотная группа	Частота в словнике	Процент	Количество словоупотр.	Процент
1	1 – 4	21565	85,0925	31876	24,8490
2	5 – 9	1905	7,5169	12438	9,6961
3	10 – 100	1769	6,9802	43257	33,7210
4	101 – 1000	98	0,3867	23653	18,4387
5	>1000	6	0,0237	17055	13,2952

И здесь видно, что наибольшую долю словника составляет низкочастотная лексика – свыше 85%; высокочастотные словоупотребления занимают лишь 7% словника, т.е. получена «традиционная» картина лексикостатистического анализа тематических текстов. Здесь терминологическое наполнение тоже весьма четко выражено, однако оно не составляет подавляющее большинство, как это было в случае с терминами по автомобилизму и технике.

5.1.8.Тексты о политике

После лексико-статистического анализа текстов был составлен частотный перечень слов, а также таблицы, отражающие статистическую структуру словника. В текстах о политике насчитывается 97709 словоупотреблений; длина словника – 11581 разное слово и словоформа. Следовательно, индекс итерации (среднее число употребительности слов) составляет – 8,4 (97709 : 11581). Коэффициент лексического разнообразия – 0,1185 (11581 : 97709).

В числе наиболее употребительных слов были: *федерация-1309; дума-1208; область-890; совет-369; год-655; закон-403; власть-243; депутат-312; собрание-186; статья-184; рубль-174; член-173; орган-170; деятельность-166; президент-272; служба-145; порядок-140; внесение-138; бюджет-137; полномочие-136 и т.д.*

Наиболее частотными были следующие имена собственные: *Россия-153; Путин-105; Москва-38; Каменская-34; Владимир-64; Константин-32; Касьянов-24; Тюмень-19; ЮКОС-17 и т.д.*

Полные данные представлены в виде частотной таблицы (см. №8).

Таблица №8

ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ О ПОЛИТИКЕ

№	Частотная группа	Частота в словнике	Процент	Количество словоуп.	Процент
1	1 – 4	8176	70,5984	16856	17,25,12
2	5 – 9	1823	15,7413	11839	12,1166
3	10 – 100	1499	12,9436	35675	36,5115
4	101 – 1000	77	0,6649	20810	21,2979
5	>1000	6	0,0518	12529	12,8228

Здесь повторяемость некоторых словоупотреблений была весьма высокой, поэтому низкочастотная группа слов составляет лишь 70,6%. Между тем высокочастотная группа несколько увеличилась – почти до 14%.

5.1.9. Анализ текстов о банковском деле

После лексикостатистического анализа установлено: а) в текстах о банковском деле насчитывается 131767 словоупотреблений; длина словника – 3631 разное

слово и словоформа. Следовательно, индекс итерации (среднее число употребительности слов) составляет – 36,3 (131767 : 3631). Коэффициент лексического разнообразия – 0,0275.

В числе наиболее употребительных слов были: *налог-4617; уплата-2360; закон-1860; федерация-1423; лицо-1711; организация-1034; налогоплательщик-916; порядок-1353; сбор-1968; имущество-1138; обязанность-735; орган-1748; статья-1648; агент-656; договор-663; законодательство-670; кодекс-1205; решение-940; счёт-594; соответствие-553 и т.д.*

Были лишь два имени собственные: *Минфин-41; Москва-40.*

Лексикостатистические данные представлены в виде результатов частотной таблицы (см.№9).

Таблица №9

СТАТИСТИЧЕСКАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ О БАНКОВСКОМ ДЕЛЕ

Ранги	Частотная группа	Частота в словнике	Процент	Количество словоупотр.	Процент
1	1 – 4	124	3,4273	490	0,2134
2	5 – 9	289	7,9878	2412	1,0504
3	10 – 100	2834	78,3306	77382	33,6996
4	101 – 1000	344	9,5080	87358	38,0441
5	>1000	27	0,7463	61981	26,9925

Своеобразие данного словника очевидна: низкочастотные словоупотребления составляют лишь 3,4%. Это самый низкий уровень низкочастотных словоупотреблений. Значит, это свидетельство о том, что в текстах имеется резвычайно большая повторяемость однотипных словоупотреблений. Это явление зафиксировано весьма большими индексами наиболее частотной лексики, которая составляет в словнике свыше 88%.

5.1.10.Тексты о спорте

Лексикостатистический анализ показал, в в текстах о спорте насчитывается 100352 словоупотребления; длина словника – 15648 разных слов и словоформ. Следовательно, индекс итерации (среднее число употребительности слов) составляет – 6,4 (100352 : 15648). Коэффициент лексического разнообразия –

0, 1559

В числе наиболее употребительных слов были: команда-618; мяч-308; газета-291; матч-727; игра-299; клуб-283; тренер-160; время-159; чемпионат-142; минута-252; счёт-248; март-132; ворота-234; игрок-123; очко-119; состав-118; год-116; хозяин-115; сезон-218; гость-215 и т.д.

Наиболее частотными были следующие имена собственные: «Динамо»-165; Россия-121; Украина-107; Москва-74; Сергей-73; «Шахтёр»-72; Александр-65; УЕФА-63; «Днепр»-111; «Торпедо»-57; «Металлург»-110; «Локомотив»-49; Владимир-43; Челси-43; «Сатурн»-42; Виктор-29; Милан-29; Англия-28; Евролига-28 и т.д.

Статистические данные приводятся в частотной таблице (см. №10).

Таблица №10

СТАТИСТИЧЕСКАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ О СПОРТЕ

Ранги	Частотная группа	Частота в словнике	Процент	Количество словоуп.	Процент
1	1 – 4	12209	78,0228	22942	22,8615
2	5 – 9	1923	12,2891	12380	12,3366
3	10 – 100	1416	9,0491	33577	33,4592
4	101 – 1000	95	0,6071	19616	19,5472
5	>1000	5	0,0320	11837	11,7955

В данном случае повторяемость слов и спортивных терминов имела обычный характер: низкочастотные словоупотребления составили большую часть словника – свыше 78%; низкочастотные словоупотребления составили лишь чуть более 10%. Тематическую окраску словника и текстов можно без труда определить по именам собственным и спортивным терминам.

5.1.11. Параметры и лексический состав сводного словника

На основе компьютерной программы “Лингвостатистика”, все тематические словники суммировались и таким образом получен сводный словарь по десяти проанализированным темам. Его параметры были следующими:

- а) всего в текстах было 1 212 390 слов и словоупотреблений;
- б) в образованном сводном словаре было 107 775 разных слов и словоформ.

Следовательно, выборка по десяти разным тематическим текстам получена статистически достоверной, как, например, в ЧСРЯ (“Частотный словарь русского языка”), ЧССРЯ (“Частотный словарь современного русского языка”) или ЧСЛЯ (“Частотный словарь литовского языка”).

По параметрическим числам можно высчитать другие показатели сводного словаря:

а) индекс итерации (отношение числа словоупотреблений в текстах к числу разных слов в словаре) равно:

$$1\ 212\ 390 : 107\ 775 = 11,25$$

б) коэффициент лексического разнообразия:

$$107\ 775 : 1\ 212\ 390 = 0,0889$$

Это значит, что каждое слово в текстах встречается в среднем через 11–12 других слов и словоформ.

Коэффициент лексического разнообразия является весьма близким к аналогичным показателям в ЧСРЯ, “Частотного словаря языка Пушкина” и т.д.

Статистическая структура сводного словаря отражается в таблице (см. №11).

Таблица №11

Ранги	Частотная группа	Частота в словаре	Относит. частота	Количество словоупотр.	Относит. частота
1	1 – 4	82129	76,2041	133176	10,9846
2	5 – 9	11706	10,8615	75715	6,2451
3	10 – 100	12564	11,6576	328541	27,0986
4	101 – 1000	1279	1,1867	306227	25,2581
5	>1000	97	0,0900	368731	30,4136

Как видно, при большой выборке наиболее значительную часть словаря тоже составляет низкочастотная лексика – свыше 76%. Около 11% составляет среднечастотная лексика, а остальную часть – около 13% – составляют высокочастотные словоупотребления.

Рассмотрим примеры высокочастотных слов сводного словаря по 10 темам:

а) ряд служебных слов, местоимений, частиц и др.: *и* – 48882, *в* – 41935, *на* – 19777, *не* – 16588, *с* – 12793, *по* – 10641, *что* – 9132, *а* – 7769, *от* – 6974, *или* – 6736, *как* – 6295, *к* – 6043, *о* – 5879, *за* – 5670, *из* – 5590, *его* – 5392, *он* – 4987, *у* – 4637, *то* – 4551, *но* – 4443 и т.д.

б) ряд высокочастотных словоупотреблений (знаменательных частей речи): *было* – 2224, *может* – 2036, *быть* – 8178, *время* – 1411, *есть* – 1360, *можно* – 1328 *очень* – 1184, *сказал* – 1164, *один* – 1026, *случай*, 1020, *меня* – 1015, *Господь* – 856, *день* – 851, *себя* – 845, *жизнь* – 717, *других* – 619, *является* – 673, *теперь* – 663, *больше* – 644, *просто* – 604, *несколько* – 635 и др.

В ряду высокочастотных примеров сводного словаря по 10 темам заметно большое влияние лексического наполнения текстов по банковскому делу. В них чрезвычайно много употребляется однотипных лексем, например: *налог, уплата, организация, налогоплательщик, сбор, органы, статья, имущество, законодательство* и т.д.

Поскольку низкочастотная лексика составляет подавляющее большинство в составе сводного словаря, то рассмотрим таблицу по каждому низкочастотному рангу (см. табл. №12)

Таблица №12

**СТРУКТУРА НИЗКОЧАСТОТНЫХ СЛОВОУПОТРЕБЛЕНИЙ
ПО СВОДНОМУ СЛОВАРЮ 10 РАЗНЫХ ТЕМ**

Частота	Повтор частот	Процент	Число словоупотреблений в текстах	Процент
4	6233	5,7833	24932	2,0564
3	7924	7,3524	23772	1,9608
2	16500	15,3097	33000	2,7219
1	51472	47,7588	51472	4,2455

Как видно из частотной структуры, самую большую долю составляют словоупотребления, которые повторялись во всех текстах лишь один раз – почти 48%.

При этом следует отметить, что с уменьшением индексов частот увеличиваются числа их повторов. Это закономерное явление низкочастотных индексов, т.е. такая тенденция наблюдается в любых исследованиях лексического наполнения разных текстов. Необходимо также подчеркнуть, что почти половина сводного словаря в десяти разных темах состоит из единичных лексем. Можно себе представить, что они в текстах могли и не появиться, о чем свидетельствуют их “пограничные” индексы частот, определённые лишь единицей повторов. Следовательно, эту часть лексем условно можно было бы отнести к пассивной

части словаря, что будет дополнительно проиллюстрировано на примерах родного (литовского) языка.

Полную частотную структуру сводного словаря в кратком виде хорошо иллюстрирует таблица данных, распределённых по группам частот (см. табл.№13).

Таблица №13

Распределение словоупотреблений в сводном словнике по группам частот

Ранги	Частотная группа	Частота в словаре	Относит. частота	Количество словоупотр.	Относит. частота
1	1 – 4	82129	76,2041	133176	10,9846
2	5 – 9	11706	10,8615	75715	6,2451
3	10 – 100	12564	11,6576	328541	27,0986
4	101 – 1000	1279	1,1867	306227	25,2581
5	>1000	97	0,0900	368731	30,4136

Как видно по результатам таблицы, наибольшую долю словаря составляют низкочастотные словоупотребления (индексы частот которых 1-4); их свыше 76%.

Высокочастотная лексика составляет около 13%. Это наиболее употребительные слова, которые как бы пронизывают все тематические выборки и составляют лексическое ядро данного словника.

Таким образом, в результате досконального лексико-статистического исследования состава десяти тематических текстов, произведена дифференциация слов по интенсивности словоупотреблений. Однако при последовательном суммировании тематических словников, дифференциация лексики меняется: тематические слова перестают быть наиболее употребительными единицами, и ядро высокочастотных слов формируется за счёт всех тематических текстов.

5.1.12. Анализ высокочастотной заимствованной лексики

Исследование лексики текстов разной тематики методами лингвистической статистики сопряжено определением ядра и периферии лексической системы (активного и пассивного словаря).

По результатам исследования мы отметили, что лексическое ядро является более устойчивым по своему составу, чем периферия.

Чтобы сказанное было очевидным в большей степени, мы поставили перед собой дополнительную цель – определить частотность заимствований по кластерному (интервальному) принципу в пределах первых пятисот наиболее употребительных знаменательных частей речи, которые являются заимствованиями из других языков (латинского, греческого, немецкого, французского и др.).

В основу дифференциации положены пометы, или указатели, приведённые в “Словаре иностранных слов”(1985г.). Заимствования рассматривались по всем десяти разным темам, объём выборок которых в среднем 100000 словоупотреблений.

Статистические результаты приводятся в виде данных иллюстративных таблиц.

ЧИСЛО ЗАИМСТВОВАНИЙ В СЛОВНИКЕ ТЕКСТОВ БИБЛИИ (Ветхого Завета)

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	-	-	-	-	-	-	-
101 – 200	-	-	-	-	-	-	-
201 – 300	-	1	-	-	-	-	1
301 – 400	-	1	-	-	-	-	1
401 – 500	-	-	-	-	-	-	-

ЧИСЛО ЗАИМСТВОВАНИЙ В СЛОВНИКЕ ТЕКСТОВ ХУДОЖЕСТВЕННОЙ ЛИТЕРАТУРЫ

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	-	-	-	-	-	-	-
101 – 200	-	-	-	-	-	-	-
201 – 300	-	2	-	-	-	-	2
301 – 400	3	1	1	-	-	-	5
401 – 500	3	-	1	-	1	-	5

ЧИСЛО ЗАИМСТВОВАНИЙ В СЛОВНИКЕ ТЕКСТОВ ПО МЕДИЦИНЕ

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	5	4	-	-	1	-	10
101 – 200	2	3	-	-	-	1	6
201 – 300	4	2	-	-	2	-	8
301 – 400	3	2	-	-	1	1	7
401 – 500	5	3	-	-	7	-	15

ЧИСЛО ЗАИМСТВОВАНИЙ В СЛОВНИКЕ ТЕКСТОВ ПО АВТОМОБИЛИЗМУ

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	1	3	-	-	-	-	4
101 – 200	1	1	-	-	1	3	6
201 – 300	4	1	-	-	1	2	8
301 – 400	3	1	-	-	1	2	7
401 – 500	4	1	-	-	3	6	14

ЧИСЛО ЗАИМСТВОВАНИЙ В СЛОВНИКЕ ТЕКСТОВ О ВОЙНЕ

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	-	-	-	-	2	-	2
101 – 200	4	2	-	-	-	3	9
201 – 300	1	1	-	1	-	-	3
301 – 400	4	5	-	-	3	-	12
401 – 500	4	4	-	-	1	-	9

ЧИСЛО ЗАИМСТВОВАНИЙ В СЛОВНИКЕ ТЕКСТОВ О ЖЕНЩИНАХ

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	-	-	-	-	-	-	-
101 – 200	-	1	-	-	1	-	2
201 – 300	1	1	-	-	1	-	3
301 – 400	-	-	-	1	2	-	3
401 – 500	4	2	-	-	-	1	7

ЧИСЛО ЗАИМСТВОВАНИЙ В СЛОВНИКЕ ТЕКСТОВ ПО ЖИВОТНОМУ МИРУ

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	-	-	-	-	-	-	-
101 – 200	-	1	-	-	-	-	1
201 – 300	1	1	-	-	-	-	2
301 – 400	-	1	-	-	-	-	1
401 – 500	1	1	-	-	-	-	2

ЧИСЛО ЗАИМСТВОВАНИЙ В СЛОВНИКЕ ТЕКСТОВ О ПОЛИТИКЕ

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	4	2	-	-	-	-	6
101 – 200	6	2	-	-	4	-	12
201 – 300	6	5	-	-	-	-	11
301 – 400	4	3	-	-	2	-	9
401 – 500	5	5	-	-	1	-	11

ЧИСЛО ЗАИМСТВОВАНИЙ В СЛОВНИКЕ ТЕКСТОВ ПО БАНКОВСКОМУ ДЕЛУ

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	5	2	-	-	1	-	8
101 – 200	8	2	-	1	2	-	13
201 – 300	2	2	-	-	3	-	7
301 – 400	5	-	-	-	2	1	8
401 – 500	4	1	-	-	-	1	6

ЧИСЛО ЗАИМСТВОВАНИЙ В СЛОВНИКЕ ТЕКСТОВ ПО СПОРТУ

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	1	-	3	-	2	-	6
101 – 200	3	-	3	-	2	-	8
201 – 300	2	-	4	-	2	-	8
301 – 400	4	-	4	1	4	-	13
401 – 500	1	-	4	-	3	1	9

**ЧИСЛО ЗАИМСТВОВАНИЙ В СВОДНОМ СЛОВНИКЕ ИНОСТРАННЫХ СЛОВ
ПО 10 ТЕМАМ
(суммарная характеристика)**

Кластер (интервал)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	16 13,11%	11 15,72%	3 15,0%	0	6 10,72%	2 20,0%	36 12,25%
101 – 200	28 22,96%	12 17,15%	3 15,0%	1 25,0%	10 17,85%	7 31,82%	61 20,75%
201 – 300	21 17,21%	16 22,85%	4 20,0%	1 25,0%	9 16,07%	2 9,09%	53 18,02%
301 – 400	26 21,32%	14 20,0%	5 25,0%	2 50,0%	15 26,78%	4 18,19%	66 22,45%
401 – 500	31 25,40%	17 24,28%	5 25,0%	0	16 28,57%	9 40,90%	78 26,53%
Суммы по кластерам	122	70	20	4	56	22	294
Процент по кластерам	41,5%	23,81%	6,8%	1,36%	19,05%	7,48%	100%

Таким образом видно, что наибольшее число заимствований отмечается в тематических текстах по медицине (25,94%) и политике (14,41%);

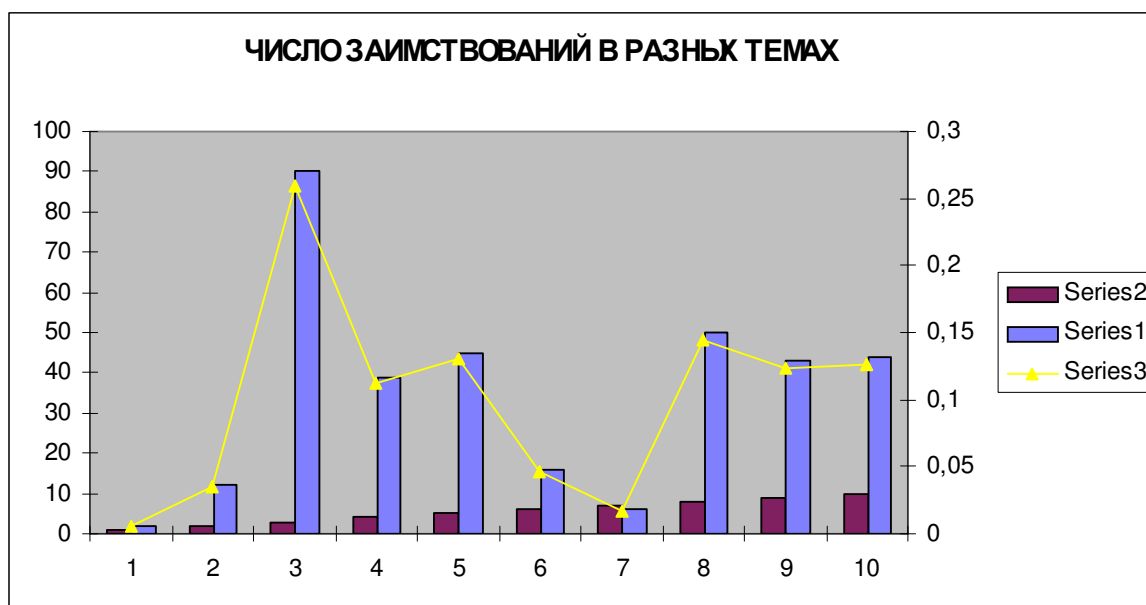
Наименьшее число заимствований – в текстах по Ветхому Завету (0,58%) и в текстах по животному миру (1,73%).

В целом статистическая структура заимствований по десяти темам выглядит следующим образом (см. табл. №14 и график №1).

СТАТИСТИЧЕСКАЯ СТРУКТУРА ЗАИМСТВОВАНИЙ ПО ДЕСЯТИ ТЕМАМ

Словники	Количество заимствований	Относительное число
1.Словник текстов Библии	2	0,005764
2.Словник текстов по художественной литературе	12	0,034582
3.Словник текстов по медицине	90	0,259366
4.Словник текстов по автомобилизму	39	0,112392
5.Словник по текстам о войне	45	0,129683
6.Словник по текстам о женщинах	16	0,04611
7.Словник текстов по животному миру	6	0,017291
8.Словник текстов по политике	50	0,144092
9.Словник текстов по банковскому делу	43	0,123919
10.Словник текстов по спорту	44	0,126801

График №1



ОБОЗНАЧЕНИЯ:

- 1.Словник текстов Библии (Ветхого завета);
- 2.Словник текстов художественной литературы;
- 3.Словник текстов по медицине;
- 4.Словник текстов по автомобилизму;
- 5.Словник по текстам о войне;
- 6.Словник по текстам о женщинах;
- 7.Словник текстов по животному миру;
- 8.Словник текстов по политике;

9.Словник текстов по банковскому делу;

10.Словник текстов по спорту.

Далее рассмотрим конкретные примеры заимствований по всем темам в пределах 500 наиболее частых словоупотреблений:

В СЛОВНИКЕ ТЕКСТОВ БИБЛИИ (ВЕТХОГО ЗАВЕТА)

Интервалы (кластеры)	Имена существительные (нарицательные)
1 – 100	-
101 – 200	-
201 – 300 (всего 1)	Фараон-(3)187*5 gr<egipt.
301 – 400 (всего 1)	Том-(1)38*130 gr.
401 – 500	-

В СЛОВНИКЕ ТЕКСТОВ ХУДОЖЕСТВЕННОЙ ЛИТЕРАТУРЫ

Интервалы (кластеры)	Имена существительные (нарицательные)
1 – 100	-
101 – 200	-
201 – 300 (всего 2)	Кино-(1)38*51 gr.; поэт-(1)33*109 gr.
301 – 400 (всего 5)	Премия-(2)57*41 lot.; момент-(1)29*181 lot.; поэт-(1)29*109 gr.; фильм-(1)29*73 angl.; история-(1)25*382 gr.
401 – 500 (всего 5)	Роман-(1)24*117 pranc.; эффект-(1)24*37 lot.; фильм-(1)22*73 angl.; автор-(1)21*202 lot.; телевизор-(1)21*17 lot.

В СЛОВНИКЕ ТЕКСТОВ ПО МЕДИЦИНЕ

Интервалы (кластеры)	Имена существительные (нарицательные)
1 –100 (всего 10)	Эндометрия-(1)- gr.; цикл-(1)164*22 gr.; функция-(1)147*27 lot.; терапия-(1)- gr.; фон-(1)116*36 pranc.; процесс-(1)111*293 lot.; препарат-(2)205*12 lot.; орган-(1)102*134 gr.; система-(1)97*289 lot.; канал-(1)95*107 lot.
101 – 200 (всего 19)	Эпителий-(1)92*- gr.; процесс-(3)233*293 lot.; эстроген-(1)85*- gr.; масса-(1)76*305 lot.; пациентка-(1)74*1 lot.; характер-(1)74*236 gr.; период-(1)71*156 gr.; секреция-(1)70*- lot.; полип-(1)68*2 gr.; псевдоэрозия-(1)66*- gr.+lot.; пролактин-(1)64*- lot.; том-(1)64*130 gr.; менструации-(1)63*- lot.; прогестерон-(1)63*- lot.; препарат-(1)61*6 lot.; форма-(1)61*456 lot.; овуляция-(1)60*- lot.; гипофиз-(1)59*-gr.; гормон-(1)58*2gr.
201 – 300 (всего 14)	Патология-(1)56*- gr.; гиперплазия-(1)54*- gr.; контрацепция-(1)53*- lot.; тема-(1)52*140 gr.; результат-(1)51*216 pranc.; терапия-(1)50*- gr.;

	гипертензия-(1)49*- gr.+lot.; метод-(2)93*170 gr.; период-(1)47*156 gr.; рецептор-(1)46*11 lot.; полип-(1)45*2 gr.; синдром-(1)45*- gr.; плазма-(1)44*- gr.; группа-(1)43*244 vok..
301 – 400 (всего 20)	Синтез-(1)41*31 gr.; диагностика-(1)40*1 gr.; дисплазия-(1)40*- lot.; мастопатия-(1)40*- gr.; миома-(1)39*- gr.; эстроген-(2)39*- gr.; андроген-(1)38*- gr.; автор-(1)37*202 lot.; симптом-(2)37*- gr.; сперматозоид-(1)37*- gr.; гиповолемиа-(1)36*- gr.; диагноз-(1)36*1 gr.; апатия-(1)35*1 gr.; метод-(1)35*170 gr.; параметр-(1)35*3 gr.; препарат-(1)35*6 lot.; норма-(1)34*63 lot.; реакция-(1)34*66 lot.; эрозия-(1)34*2 lot.; процесс-(1)33*293 lot.
401 – 500 (всего 27)	Эрозия-(1)32*2 lot.; контроль-(1)31*80 pranc.; операция-(1)31*94 lot.; организм-(2)31*70 lot.; терапия-(1)31*- gr.; фаза-(1)31*11 gr.; гиперпролактинемия-(1)30*- gr.; диагностика-(1)30*1 gr.; псевдоэрозия-(1)30*- gr.; секреция-(1)30*10 lot.; тестостерон-(1)30*- lot.; группа-(1)29*244 vok.; доза-(2)29*2 gr.; мина-(1)29*29 pranc.; плазма-(1)29*20 gr.; эндометриоз-(1)29*- gr.; пост менопауза-(1)28*- lot.; режим-(1)28*84 pranc.; тип-(1)28*89 gr.; эндометрия-(2)28*- gr.; ангиотензиноген-(1)27*- gr.; анемия-(1)27*- gr.; метод-(1)27*170 gr.; прогестин-(1)27*- lot.; фактор-(1)27*51 lot.; характер-(1)27*236 gr.

В СЛОВНИКЕ ТЕКСТОВ ПО АВТОМОБИЛИЗМУ

Интервалы (кластеры)	Имена существительные (нарицательные)
1 – 100 (всего 4)	Федерация-(1)1502*11 lot.; автомобиль-(3)1796*279 gr.+lot.; машина-(2)1016*1056 gr.+lot.; том-(1)586*130 gr.
101 – 200 (всего 6)	Территория-(1)338*95 lot.; машина-(1)334*523 pranc.+lot.; полис-(2)318*- pranc.<lot.; орган-(2)260*134 gr.; тариф-(1)254*5 pranc.<arab.; компания-(1)230*73 pranc.
201 – 300 (всего 8)	Коэффициент-(1)206*16 lot.; организация-(1)204*372 pranc.<lot.; субъект-(1)196*3 lot.; эксплуатация-(1)182*87 pranc.; сумма-(1)180*62 lot.; орган-(1)174*134 gr.; автомобиль-(1)170*93 gr.+lot.; премия-(1)168*41 lot.
301 – 400 (всего 7)	Информация-(1)156*57 lot.; авария-(1)152*16 it.<arab.; ремонт-(1)152*60 pranc.; автомобиль-(1)132*93 gr.+lot.; коэффициент-(1)132*16 lot.; форма-(1)126*456 lot.; период-(1)124*156 gr.
401 – 500 (всего 14)	Сумма-(2)122*62 lot.; документ-(2)114*103 lot.; пункт-(2)114*80 lot.; автомобиль-(2)224*93 gr.+lot.; акт-(1)112*62 gr.+lot.; компания-(2)224*73 pranc.; машина-(2)224*523 pranc.<lot.; реализация-(1)112*13 pranc.; категория-(1)110*37 gr.; контроль-(1)110*80 pranc.; результат-(1)110*216 pranc.<lot.; модель-(1)108*44 pranc.<it.<lot.; тариф-(1)106*5 pranc.<arab.; экспертиза-(1)106*1 lot.

В СЛОВНИКЕ ТЕКСТОВ О ВОЙНЕ

Интервалы (кластеры)	Имена существительные (нарицательные)
1 – 100 (всего 11)	Армия-(2)294*557 pranc.<lot.; террор-(1)98*16 pranc.<lot.
101 – 200 (всего 6)	Террор-(1)73*16 pranc.<lot.; информация-(1)70*57 lot.; организация-(1)59*372 pranc.<gr.; анархия-(1)45*21 gr.; конфликт-(1)44*22 lot.; территория-(1)43*95 lot.
201 – 300	Религия-(2)76*33 pranc.<lot.; армия-(1)41*557 pranc.<lot.; группа-

(всего 7)	(2)36*244 vok.; корреспондент-(1)36*160 vok.<it.; президент-(1)35*100 lot.; фронт-(2)69*176 pranc.<lot.; партия-(1)34*532 lot.
301 – 400 (всего 12)	Позиция-(1)32*129 lot.; стадия-(1)31*31 gr.; журналист-(1)30*82 pranc.; оператор-(1)30*7 lot.; террор-(1)30*16 pranc.<lot.; мораль-(2)29*88 pranc.<lot.; программа-(1)29*179 gr.; анархия-(1)28*21 gr.; генерал-(1)28*327 lot.; политик-(1)27*38 gr.; проблема-(1)27*257 gr.; центр-(1)27*161 lot.
401 – 500 (всего 9)	Организация-(1)26*372 pranc.<gr.; процесс-(1)25*293 lot.; телекомпания-(1)25*1 gr.; интерес-(1)24*325 vok.<lot.; революция-(1)24*407 lot.; информация-(1)23*57 lot.; система-(1)23*289 gr.; характер-(1)23*236 gr.; история-(1)22*382 gr.

В СЛОВНИКЕ ТЕКСТОВ О ЖЕНЩИНАХ

Интервалы (кластеры)	Имена существительные (нарицательные)
1 – 100	-
101 – 200 (всего 2)	История-(1)74*382 gr.; организация-(1)56*372 pranc.<gr.
201 – 300 (всего 3)	Республика-(1)45*244 lot.; реклама-(1)41*32 pranc.<lot.; проблема-(1)34*275 gr.
301 – 400 (всего 4)	Организация-(1)32*372 pranc.<gr.; интерес-(1)31*325 vok.<lot.; период-(1)26*156 gr.; реклама-(1)26*pranc.<lot.
401 – 500 (всего 7)	Реклама-(1)25*32 pranc.<lot.; форма-(1)22*456 lot.; анализ-(1)21*115 gr.; культура-(1)21*269 lot.; республика-(1)244 lot.; система-(1)21*289 gr.; автор-(1)20*202 lot.

В СЛОВНИКЕ ТЕКСТОВ ПО ЖИВОТНОМУ МИРУ

Интервалы (кластеры)	Имена существительные (нарицательные)
1 – 100	-
101 – 200 (всего 1)	Период-(1)71*156 gr.
201 – 300 (всего 2)	Том-(1)63*130 gr. масса-(1)58*309 lot.
301 – 400 (всего 1)	Метр-(1)39*171 gr.
401 – 500 (всего 2)	Километр-(1)36*251 gr.; территория-(1)35*95 lot.

В СЛОВНИКЕ ТЕКСТОВ О ПОЛИТИКЕ

Интервалы (кластеры)	Имена существительные (нарицательные)
1 – 100 (всего 6)	Федерация-(1)1309*11 lot.; депутат-(2)191*69 lot.; орган-(1)170*134 gr.; президент-(2)150*100 lot.; том-(1)122*130 gr.; территория-(1)94*95 lot.
101 – 200 (всего 12)	Депутат-(1)89*69 lot.; субъект-(1)86*3 lot.; фонд-(1)83*36 pranc.<lot.; администрация-(1)82*19 lot.; реализация-(1)64*13 pranc.; политик-(1)63*38 gr.; проект-(2)60*76 lot.; информация-(1)57*57 lot.; субъект-

	(1)57*3 lot.; организация-(1)56*372 pranc.<gr.; орган-(2)110*134 gr.; инициатива-(1)53*66 pranc.<lot.
201 – 300 (всего 12)	Культура-(1)52*269lot.; депутат-(1)51*69 lot.; продукция-(1)50*161 lot.; комиссия-(1)49*110 lot.; орган-(1)49*134 gr.; программа-(1)49*179 gr.; комплекс-(1)47*33 lot.; проблема-(1)45*257 gr.; аппарат-(1)43*128 lot.; стратегия-(1)41*9 gr.; период-(1)40*156 gr.; президент-(1)39*100 lot.
301 – 400 (всего 9)	Сумма-(1)38*62 lot.; депутат-(1)37*69 lot.; история-(1)37*382 gr.; конституция-(1)37*27 lot.; институт-(1)35*274 lot.; организация-(1)34*372 pranc.<gr.; комитет-(1)33*160 pranc.; система-(1)32*289 gr.; тема-(1)32*140 gr.
401 – 500 (всего 11)	Объект-(1)31*73 lot.; орган-(1)30*134 gr.; программа-(2)60*179 gr.; депутат-(1)29*69 lot.; контроль-(1)29*80 pranc. система-(1)29*289 gr.; комиссия-(1)28*110 lot.; акт-(1)27*62 lot.; экономика-(1)27*135 gr.; база-(1)26*249 gr.; президент-(1)26*100 lot.

В СЛОВНИКЕ ТЕКСТОВ ПО БАНКОВСКОМУ ДЕЛУ

Интервалы (кластеры)	Имена существительные (нарицательные)
1 – 100 (всего 8)	Федерация-(1)1423*11 lot.; организация-(1)1034*372 pranc.<gr.; орган-(4)2080*134 gr.; агент-(1)656*50 lot.; кодекс-(2)1205*23 lot.; кредит-(1)546*6 lot.; сумма-(3)1268*62 lot..
101 – 200 (всего 14)	Консультант-(1)322*3 lot.; том-(1)320*130 gr.; кредит-(2)847*6 lot. орган-(2)492*134 gr.; регистрация-(1)269*4 lot.; пункт-(1)257*80 lot.; сумма-(1)248*62 lot.; документы-(2)243*103 lot.; фонд-(1)220*36 pranc.<lot.; организация-(1)208*372 pranc.<gr.; полиция-(1)204*59 vok.<gr.; документ-(1)194*103 lot.; процент-(1)187*172 lot.
201 – 300 (всего 7)	Момент-(2)310*181 lot.; система-(1)180*289 gr.; реорганизация-(1)174*- pranc.; форма-(1)167*456 lot.; организация-(1)166*372 pranc.<gr.; период-(1)149*156 gr.; фонд-(1)131*36 pranc.<lot.
301 – 400 (всего 8)	Регистрация-(1)120*4 lot.; пункт-(1)116*80 lot.; копия-(1)114*14 lot.; документ-(2)208*103 lot.; банк-(2)208*39 it.; организация-(1)102*372 pranc.<gr.; агент-(1)100*50 lot.; реализация-(1)100*13 pranc.
401 – 500 (всего 6)	Акт-(1)95*62 lot.; база-(1)91*249 gr.; банк-(2)177*39 it.; рефинансирование-(1)88*-lot.; форма-(1)87*456 lot.; сумма-(1)80*62 lot.

В СЛОВНИКЕ ТЕКСТОВ О СПОРТЕ

Интервалы (кластеры)	Имена существительные (нарицательные)
1 – 100 (всего 6)	Команда-(3)618*180 pranc.<lot.; клуб-(2)283*121 angl.; тренер-(1)160*31 angl.; чемпионат-(1)142*31 angl.; минута-(2)252*732 lot.; сезон-(2)218*76 pranc.
101 – 200 (всего 8)	Команда-(3)235*180 pranc.<lot.; чемпионат-(1)85*31 angl.; лига-(1)82*3 pranc.<lot.; тренер-(1)160*31 angl.; минута-(1)141*366 lot.; момент-(1)69*181 lot.; тайм-(1)68*- angl.; факт-(1)59*225 lot.
201 – 300 (всего 8)	Чемпион-(1)57*60 angl.; сайт-(1)52*- angl.; команда-(1)51*180 pranc.<lot.; клуб-(1)45*121 angl.; тайм-(2)101*- angl.; результат-(1)42*216 lot.; тур-(1)42*13 pranc.; серия-(1)40*46 lot..

301 – 400 (всего 13)	Лига-(1)38*3 pranc.<lot.; клуб-(1)37*121 angl.; туп-(1)37*13 pranc.; линия-(1)36*196 lot.; метр-(1)36*171 pranc.<gr.; серия-(1)36*46 lot.; атака-(2)35*63 pranc.; контракт-(1)35*5lot.; спорт-(1)34*73 angl.; турнир-(1)34*35 vok.<pranc.; момент-(1)32*181 lot.; тренер-(1)31*31 angl.; футбол-(2)31*28angl.
401 – 500 (всего 9)	Капитан-(1)30*352 pranc.<lot.; футбол-(1)30*28 angl.; чемпионат-(1)30*31 angl.; президент-(1)28*100 lot.; чемпион-(1)27*60 angl.; клуб-(1)26*121 angl.; финал-(1)26*8 it.<lot.; шанс-(1)26*13 pranc.; арбитр-(1)25*4 pranc.<lot.

6.ЛЕКСИКОСТАТИСТИЧЕСКИЙ АНАЛИЗ РАЗНЫХ ТЕКСТОВ НА ЛИТОВСКОМ ЯЗЫКЕ

В целях сопоставительного анализа лексического наполнения текстов разной тематики был выполнен ряд исследований на родном (литовском) языке.

Напомним ряд названий тематических выборок;

Это были тексты Библии, художественной литературы, разных областей науки; тексты по технике и автомобилизму; тексты о войне, женщинах, семье, политике, банковском деле, спорте.

Следовательно, методами лингвистической статистики проверялось адекватность или неадекватность лексикостатистических результатов о функционировании слов в русском и литовском языках. Поначалу будут представлены данные об исследовании тематического наполнения разных текстов с иллюстрациями словоупотреблений по их частотам; в другом разделе будут рассмотрены результаты сопоставления.

6.1. Лингвостатистическая характеристика тематических текстов

В данном разделе будут представлены результаты о наиболее частых и редких (низкочастотных) словоупотреблениях, а также статистические таблицы по частотным группам, чтобы таким способом можно было выявить сходства или различия статистической структуры частотных словников.

6.1.1.Тексты БИБЛИИ (на литовском языке)

Всего в указанных текстах было 105272 словоупотребления. В составленном алфавитно-частотном словнике значилось 16426 разных лексем. В числе наиболее частотных были:

а) из служебного ряда слов: *ir(u)* – 4595; *iš(из)*-1398; *į(в/на)*-1006; *su(c)*-903; *kaip(как)*-757; *kad(что)*-727; *o(a)*-597; *ar(ули)*-551; *bet(но)*-533; *ant(на)*-516; *kai(когда)*-500; *už(за)*-490; *nes(потому)*-453; *nuo(от)*-435; *tada(тогда)*-582; *jei(если)*-393; *prie(y)*-363; *tik(только)*-276; *ka(что)*-273; *ras(к)*-233 и т.д.

б) из числа знаменательных частей речи: *savo(свой)*-1444; *jis(он)*-1269; *buvo(был)*-900; *Viešpats(Господь)*-881; *ура(есть)*-806; *tavo(твой)*-682; *jie(они)*-648; *aš(я)*-629; *mano(мой)*-582; *Dievas(Бог)*-515; *tarė(сказал)*-364; *tu(ты)*-364; *ji(она)*-309; *diena(день)*-207; *visi(все)*-185; *žemė(земля)*-181; *kraštas(край)*-174; *atsakė(ответил)*-173; *sūnus(сын)*-168; *kunigas(ксендз)*-163 и т.д.

в) из ряда имён собственных: *Mozė(Моисей)*-(2)494; *Egiptas(Египет)*-210; *Jokūbas(Яков)*-107; *Izraelis(Израиль)*-88; *Juozapas(Иосиф)*-87; *Aaronas(Аарон)*-147; *Abraomas(Авраам)*-63; *Jobas(Иов)*-47; *Kanaanas(Канаан)*-44; *Izaokas(Исаак)*-41 и т.д.

Таким образом, высокочастотные слова действительно отражают тематическую сущность данных текстов.

Среди низкочастотных слов отметим такие примеры: *užsipilti(залить)*-1; *užsukti(завернуть)*-1; *užtikti(застать)*-1; *užverti(оградить)*-1; *vaikinas(юноша)*-1; *vaisingas(плодоносный)*-1; *valgis(еда)*-1; *valyti(чистить)*-1; *vargti(томиться)*; *varnas(ворон)*-1; *varžytis(соостязаться)*-1; *vasara(лето)*-1; *vedlys(проводник)*-1; *veidrodis(зеркало)*-1; *vėliava(флаг)*-1; *vėlyvas(поздний)*-1; *versmė(источник)*-1; *vidus(внутренность)*-1; *vieningai(единодушно)*-1; *vienintelis(единственный)*-1; *viešpatauti(господствовать)*-1 и т.д.

Следовательно, в числе низкочастотных словоупотреблений также имеется немало тематических признаков изучаемых текстов.

Далее рассмотрим совокупную статистическую структуру словника по группам частот (см. табл.№14):

СТАТИСТИЧЕСКАЯ СТРУКТУРА СЛОВНИКА ПО ГРУППАМ ЧАСТОТ

Ранги	Частотная группа	Частота в словнике	Относит. частота	Колич. словоупотр.	Относит. частота
1	1 – 4	13628	82,9660	20769	19,7289
2	5 – 9	1395	8,4926	8928	8,4809
3	10 – 100	1276	7,7682	31828	30,2341
4	101 – 1000	122	0,7427	34035	32,3305
5	>1000	5	0,0304	9712	9,2256

Наибольшую долю словника составляют низкочастотные словоупотребления – почти 83%; высокочастотные словоупотребления составляют лишь 7,5%.

6.1.2.Тексты художественной литературы (на литовском языке)

Всего в указанных текстах было 101123 словоупотребления. В составленном алфавитно-частотном словнике значилось 27372 разные лексемы. В числе наиболее частотных были:

а) из служебного ряда слов: *ir(u)*-4116, *i(v/na)*-1389, *iš(из)*-701, *kaip(как)*-684, *o(a)*-684, *kad(что,чтобы)*-625, *bet(но)*-601, *su(c)*-493, *ne(нет)*-458, *tik(только)*-426, *ant(на)*-417, *ar(a)*-342, *prie(y)*-287, *kai(когда)*-286, *nuo(от)*-271, *už(за)*-254, *po(под)*-243, *per(через)*-228, *apie(об)*-211 и т.д. ;

б) из числа знаменательных частей речи: *jis(он)*-1050, *ji(она)*-879, *aš(я)*-856, *būti(быть)*-680, *jie(они)*-285, *tarti(сказать)*-144, *staiga(вдруг)*-131, *tėvas(отец)*-125, *galva(голова)*-120, *žmogus(человек)*-119, *reikia(надо)*-109, *turėti(иметь)*-106, *daug(много)*-103, *imti(брать)*-97, *gerai(хорошо)*-95, *ilgai(долго)*-91, *sakyti(говорить)*-91 и т.д.;

в) из ряда имён собственных: *Garšva*-290, *Stanley*-79, *Antanas*-72, *Elena*-66, *Ženia*-32, *Aldona*-26, *Aniulis*-26, *Anupras*-25, *Grigas*-23, *Konstancija*-23, *Laukaitis*-23 и т.д.

Высокочастотные слова, как уже говорилось, определяют тематическую сущность данных текстов.

Среди низкочастотных слов отметим такие примеры: *vizija*, *vogčiom*, *vokas*, *vokietė*, *vokietukai*, *volungė*, *vonia*, *voverė*, *zigzagas*, *zuikis*, *zvimbti*, *žadinti*, *žaibas*,

žaidimas, žaislas, žaizda, žaltys, žalias, žalingas, žalsvas, žandai, žandarai, žąsis, žavėti, žavu, žegnotis, žemelė, žemuogė, žėrutis, žiauna, žiauriai, žibalas, žibintas, žiburys, žydelkaitė, žydintis, žydukas, žilas, žymus, žolė и т.д.

Как видно, в числе низкочастотных словоупотреблений также имеется немало тематических признаков исследуемых текстов.

Далее рассмотрим совокупную статистическую структуру словника по группам частот (см. табл. №15):

Таблица №15

СТРУКТУРА СЛОВНИКА ПО ТЕКСТАМ ХУДОЖЕСТВЕННОЙ ЛИТЕРАТУРЫ

Ранги	Частотная группа	Частота в словнике	Относит. частота	Количество словоупотр.	Относит. частота
1	1 – 4	24477	89,4235	35877	35,4786
2	5 – 9	1653	6,0390	10511	10,3943
3	10 – 100	1156	4,2233	26292	26,0000
4	101 – 1000	83	0,3032	21888	21,6449
5	>1000	3	0,0110	6555	6,4822

Как видно, в данном словнике тоже преобладают низкочастотные словоупотребления; они составляют свыше 89%. Высокочастотная доля словника составляет лишь 4,5%.

6.1.3. Тексты по медицине (на литовском языке)

Всего в указанных текстах было 98399 словоупотребления. В составленном алфавитно-частотном словнике значилось 24449 разных лексем. В числе наиболее частотных были:

а) из служебного ряда слов: *ir(u)*-3188, *kad(что)*-763, *ar(a)*-678, *i(в)*-653, *su(c)*-545, *nuo(от)*-496, *iš(из)*-460, *kai(когда)*-342, *po(под)*-295, *apie(об)*-292, *dėl(на счёт)*-286, *per(через)*-278, *iki(до)*-248, *jei(если)*-223, *bet(но)*-219, *be(без)*-163, *dar(ещё)*-161, *jeigu(если)*-123, *nors(хотя)*-123, *nes(потому)*-112, *net(даже)*-112, *prie(y)*-76 и т.д.;

б) из числа знаменательных частей речи: *galėti(мочь)*-558, *tai(это)*-449, *būti(быть)*-434, *jie(они)*-258, *labai(очень)*-239, *metas(нора)*-236, *jis(он)*-221, *liga(болезнь)*-214, *smegenys(мозг)*-192, *turėti(иметь)*-181, *metai(год)*-178, *gydymas(лечение)*-158, *dažnai(часто)*-155, *vaikas(ребёнок)*-130, *amžius(возраст)*-116, *gydyti(лечить)*-116, *alkoholis(алкоголь)*-109, *vaistas(лекарство)*-106,

ligonis(больной)-100, gyvenimas(жизнь)-92, darbas(работа)-94, gerti(пить)-93, vartoti(принимать)-89, simptomas(симптом)-85 и т.д.;

в) из ряда имён собственных: *Lietuva-128, Vilnius-43* и др.

Высокочастотные слова определяют тематическую сущность данных текстов.

Среди низкочастотных слов отметим такие примеры: *vyrauti, virintas, vyriškas, viršininkas, viršelis, viršyti, virtuali, virusologinė, visagalis, visaip, visata, visavertis, visiems, visuma, visuminis, visuomeninė, visuotinė, visur, viso, vitaminas, vytimas, vizija, vizitas, vizualusis, vožtuvas, voras, zona, žagsėjimas, žaisti, žaizda, žala, žaliai, žanras, žalojantis, žarna, želmuo, žemė, žemesnis, ženklai, žmonija, žudikas, žurnalas, žurnalistė, žvynelinė* и т.д.

Следовательно, в числе низкочастотных словоупотреблений также имеется немало признаков о тематике изучаемых текстов.

Далее рассмотрим совокупную статистическую структуру словника по группам частот (см. табл.№16):

Таблица №16

ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ ПО МЕДИЦИНЕ

Ранги	Частотная группа	Частота в словнике	Относит. частота	Колич. Словоуп.	Относит. частота
1	1 – 4	21164	86,5639	31118	31,6243
2	5 – 9	1783	7,2927	11488	11,6749
3	10 – 100	1418	5,7998	32928	33,4638
4	101 – 1000	83	0,3395	19677	19,9972
5	>1000	1	0,0041	3188	3,2399

Как видно по данным таблицы, низкочастотные словоупотребления составляют подавляющее большинство словника; высокочастотные словоупотребления составляют лишь 6,1%.

6.1.4. Тексты по автомобилизму (на литовском языке)

Всего в указанных текстах было 243236 словоупотребления. В составленном алфавитно-частотном словнике значилось 9645 разных лексем. В числе наиболее частотных были:

а) из служебного ряда слов: *ir(u)-7547, į(v/на)-1771, su(c)-1704, tačiau(но)-1674, kad(что,чтобы)-1580, iš(из)-1256, nuo(от)-1236, ne(не)-1014, tik(только)-997, per(через)-557, apie(про)-512, iki(до)-497, net(даже)-490, ant(на)-421,*

dél(паду)-411, bet(но)-406, po(под)-390, prie(у)-384, už(за)-357, nors(хотя)-325, kol(пока)-313 и т.д.;

б) из числа знаменательных частей речи: *automobilis(автомобиль)-1864, būti(быть)-1220, turėti(иметь)-661, metai(год)-648, labai(очень)-519, jis(он)-517, galėti(мочь)-506, lenktynės(гонки)-445, visi(все)-386, beveik(почти)-365, kartas(раз)-338, vairas(руль)-330, gamintojas(производитель)-297, kėbulas(кузов)-295, sistema(система)-281, valdymas(управление)-268, daugiau(больше)-248, atrodyti(выглядеть)-246, sukurti(создать)-245* и т.д.;

в) из ряда имён собственных: *Dakaras-316, Ferrari-299, Rols Royse-26), Toyota-259, Mazda-289, BMW-251, Volkswagen-241, Pajero-236, Bentley-232, Lietuva-230, Phantom-221, Suzuki-182, Subaru-181, Sabine-171, Ford-166* и т.д.

Среди низкочастотных слов отметим такие примеры: *nešiojamas, neteisingai, neterminuotas, netvarkingai, numatyti, nuorašas, nuosavybė, nuostatai, nustatymas, organizacija, originalas, padalinys, pakaba, paketas, palaimintas, panaikintas, pasirašyti, pasiūlyti, paskirti, pastatyti, raudona, redakcija, specializuotas, stažas, standartas, šildymas, šviestuvai, teisė, testas, triženklis, viešas, žalia, žymuo, žinios, žurnalai* и т.д.

Далее рассмотрим статистическую структуру тематического словника по группам частот (см. табл.№17):

Таблица №1

СТАТИСТИЧЕСКАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ ПО АВТОМОБИЛИЗМУ

Ранги	Частотная группа	Частота в словнике	Относит. частота	Количество словоупотр.	Относит. частота
1	1 – 4	2171	22,5091	4212	1,7317
2	5 – 9	1639	16,9933	12093	4,9717
3	10 – 100	5499	57,0140	126674	52,0786
4	101 – 1000	323	3,3489	76314	31,3745
5	>1000	13	0,1348	23943	9,8435

Как и в текстах на русском языке, в данном тематическом словнике мы видим аналогичный результат: низкочастотная доля словоупотреблений составляет 22,5%, а высокочастотные – свыше 60%. Таким образом, подтверждается тезис о том, что в данной тематике текстов лексическое наполнение является весьма однообразным, так как в них употребляется чрезвычайно много одинаковых терминов.

6.1.5.Тексты о войне (на литовском языке)

Всего в указанных текстах было 98247 словоупотреблений. В составленном алфавитно-частотном словнике значилось 17253 разные лексемы. В числе наиболее частотных были:

а) из служебного ряда слов: *ir(u)*-3267, *į(в/на)*-1167, *ar(a)*-857, *su(c)*-579, *kad(что,чтобы)*-524, *iš(из)*-516, *dėl(паду)*-362, *nuo(от)*-342, *arba(или)*-270, *iki(до)*-257, *tačiau(но)*-230, *per(через)*-218, *apie(про)*-201, *už(за)*-183, *po(под)*-173, *bet(но)*-171, *jeigu(если)*-137, *be(без)*-136 и т.д.;

б) из числа знаменательных частей речи: *karas(война)*-1147, *apsauga(охрана)*-791, *tarnyba(служба)*-776, *kraštas(край)*-751, *būti(быть)*-617, *valstybė(государство)*-365, *galėti(мочь)*-331, *respublika(республика)*-298, *tai(это)*-291, *straipsnis(статья)*-290, *karys(воин)*-282, *kariuomenė(войско)*-245, *įstatymas(закон)*-226, *metas(пора)*-219, *partizanas(партизан)*-178, *tvarka(порядок)*-175, *rajėgos(силы)*-172, *ministras(министр)*-147, *komisija(комиссия)*-140 и т.д.;

в) из ряда имён собственных: *Lietuva*-751, *NATO*-227, *Europa*-131, *JAV*-117, *Rusija*-99, *ES*-88, *Kęstutis*-67 и т.д.

Среди низкочастотных слов отметим такие примеры: *vėlyvas, vengras, verbuoti, versija, verslas, versti, vėtra, viceministras, vienetas, vieni, vienodas, viešpatauti, vietinis, vykdytojas, vynas, viršus, vonia, zona, žadėti, žaizda, žalias, žemė, žengti, ženklas, židinys, žygis, žymėti, žinia, žiniasklaida, žmonija, žolė, žodis, žudymas, žurnalas, žuvusieji* и т.д.

Следовательно, в числе низкочастотных словоупотреблений также имеется немало признаков о тематике изучаемых текстов.

Далее рассмотрим статистическую структуру данного тематического словника по группам частот (см. табл. №18):

Таблица №18

СТАТИСТИЧЕСКАЯ СТРУКТУРА ТЕМАТИЧЕСКОГО СЛОВНИКА О ВОЙНЕ

Ранги	Частотная группа	Частота в словнике	Относит. частота	Количество словоупотр.	Относит. частота
1	1 – 4	14096	81,7017	22900	23,3086
2	5 – 9	1612	9,3433	10560	10,7484
3	10 – 100	1442	8,3580	34188	34,7980
4	101 – 1000	100	0,5796	25018	25,4644
5	>1000	3	0,0174	5581	5,6806

Итак, лексическое наполнение данной темы образует «традиционную пирамиду» словника, в котором наибольшую долю занимают низкочастотные словоупотребления – почти 82%; высокочастотные словоупотребления составляют лишь около 9%.

6.1.6. Тексты о женщинах (на литовском языке)

Всего в указанных тематических текстах было 100959 словоупотреблений. В составленном алфавитно-частотном словнике значилось 18253 разные лексемы. Наиболее частотными были:

а) из служебного ряда слов: *ir(u)*-3268, *kad(что,чтобы)*-1403, *su(c)*-768, *į(в/на)*-682, *iš(из)*-639, *kaip(как)*-491, *jei(если)*-3646 *dėl(раду)*-358, *nuo(от)*-354, *bet(но)*-350, *per(через)*-332, *prieš(против)*-327, *po(под)*-246, *nes(потому)*-202, *iki(до)*-185, *arba(или)*-174, *iki(до)*-185, *už(за)*-155, *prie(y)*-150 и т.д.; эти слова не раскрывают какую-либо тематику;

б) из числа знаменательных частей речи: *galėti(мочь)*-402, *jis(он)*-392, *aš(я)*-311, *moteris(женщина)*-310, *būti(быть)*-306, *turėti(иметь)*-272, *ji(она)*-235, *metai(год)*-210, *labai(очень)*-171, *gimda(матка)*-159, *Dievas(Бог)*-139, *galima(можно)*-139, *reikia(надо)*-136, *dabar(сейчас)*-134, *vaikas(ребёнок)*-132, *norėti(хотеть)*-130, *metas(пора)*-124, *seksas(секс)*-123, *žmogus(человек)*-116, *gyvenimas(жизнь)*-112, *meilė(любовь)*-109, *šeima(семья)*-99 и т.д.; Можно утверждать, что данный ряд примеров весьма четко обозначает тематическую направленность текстов;

в) из ряда имён собственных: *Lietuva*-74.

Среди низкочастотных слов отметим такие примеры: *vainikėlis*, *vairuoti*, *vaisiai*, *vaisingas*, *vaisingumas*, *vaistas*, *vakar*, *valgymas*, *valanda*, *valgis*, *valia*, *vamzdelis*, *varginti*, *vargti*, *varikozė*, *vartojimas*, *vesti*, *veiksmas*, *velnio*, *venerinė*, *vena*, *vergė*, *verksmas*, *versija*, *verslininkė*, *vėžys*, *vidinė*, *vidus*, *vienodas*, *vietinis*, *vilkas*, *viltis*, *žaginti*, *žaisti*, *žalingas*, *žarna*, *žeidžia*, *žiauri*, *žydas*, *žiedas*, *žievė*, *žymi*, *žindanti*, *žinyba*, *žinoti*, *žmogeliukas*, *žmogždudys*, *žudyti*, *žurnalas*, *žvaigždė*, *žvaigždėlė* и т.д.

Следовательно, в числе низкочастотных словоупотреблений тоже есть целый ряд тематических признаков изучаемых текстов.

Далее рассмотрим статистическую структуру тематического словника по группам частот (см. табл. №19):

ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ О ЖЕНЩИНАХ

Ранги	Частотная группа	Частота в словнике	Относит. частота	Количество словоупотр.	Относит. частота
1	1 – 4	14725	80,6717	26571	26,3186
2	5 – 9	1969	10,7873	12871	12,7487
3	10 – 100	1457	7,9822	33195	32,8797
4	101 – 1000	100	0,5479	23651	23,4263
5	>1000	2	0,0110	4671	4,6266

Как и предыдущий словник, так и данный состоит из неимоверно большой доли низкочастотных словоупотреблений (почти 81%) и относительно незначительного количества высокочастотных слов (чуть более 8%).

6.1.7. Тексты о сейме (на литовском языке)

Всего в указанных текстах было 119157 словоупотребления. В составленном алфавитно-частотном словнике значилось 16 588 разных лексем. В числе наиболее частотных были:

а) из служебного ряда слов: *ir(u)*-3598, *kad(что,чтобы)*-1912, *ar(a/ли)*-926, *iš(из)*-867, *i(в/на)*-777, *su(c)*-556, *bet(но)*-501, *ne(не)*-464, *jeigu(если)*-431, *už(за)*-426, *kai(когда)*-364, *po(под)*-331, *apie(об)*-293, *tik(только)*-280, *todėl(потому)*-271, *iki(до)*-218, *nors(хотя)*-216, *be(без)*-195, *arba(или)*-187, *per(через)*-185, *nuo(от)*-170 и т.д.;

б) из числа знаменательных частей речи: *yra(есть)*-1738, *pirmininkas(председатель)*-932, *aš(я)*-908, *ištatymas(закон)*-871, *seimas(сейм)*-753, *kolega(коллега)*-743, *būti(быть)*-557, *dabar(сейчас)*-456, *labai(очень)*-432, *valstybė(государство)*-394, *gerbiamasis(уважаемый)*-392, *straipsnis(статья)*-388, *projektas(проект)*-292, *manyti(предполагать)*-282, *galėti(мочь)*-224, *narys(член)*-222, *komitetas(комитет)*-220, *darbas(работа)*-218, *reikia(надо)*-208, *šiandien(сегодня)*-189, *norėti(хотеть)*-188, *ponas(господин)*-184 и т.д.; на наш взгляд, данный ряд слов тоже обладает яркими тематическими признаками;

в) в числе имён собственных были: *Lietuva*-434, *Europa*-257, *Jackūnas*-246, *Listavičius*-80, *Andriukaitis*-68, *Razma*-51 и т.д.; имена собственные наиболее ярко определяют тематику текстов.

Таким образом, в целом высокочастотные слова в наибольшей степени определяют тематическую сущность данных текстов.

Среди низкочастотных слов отметим такие примеры: *veidmainystė, veikėjas, veiksmas, vekseliai, vėliava, verkti, versija, vertė, vesti, vežėjas, viceministrė, vidurkis, vidutinis, vienybė, vienišas, vietovė, vykdytojas, vilkikas, viltis, vinis, viršyti, viršus, visapusiškas, visokeriopai, visuotinis, vitaminai, viza, vokas, voverė, zakristijonas, žadėti, žaidimas, žaislinis, žaliava, žemesnis, žemynas, ženklas, žinutė, žmoniškas, žurnalistas, žvejams, žvelgti, žvilgsnis, žvilgtelti* и т.д.

Следовательно, в числе низкочастотных словоупотреблений тематику изучаемых текстов угадать значительно труднее, хотя некоторые ее признаки обнаружить вполне возможно.

Далее рассмотрим статистическую структуру словника по группам частот (см. табл.№20):

Таблица №20

СТАТИСТИЧЕСКАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ О СЕЙМЕ

Ранги	Частотная группа	Частота в словнике	Относит. частота	Количество Словоупотр.	Относит. частота
1	1 – 4	13486	81,2997	20598	17,2864
2	5 – 9	1509	9,0969	9791	8,2169
3	10 – 100	1418	8,5483	36469	30,6058
4	101 – 1000	170	1,0248	42850	35,9610
5	>1000	5	0,0301	9449	7,9299

Как показывают результаты данной таблицы, статистическая структура данных текстов ничем не отличается от многих других, не обладает каким-то своеобразием. Наибольшее число словоупотреблений относится к низкочастотной группе (свыше 81%), а высокочастотные словоупотребления составляют самую незначительную долю словника – лишь 8,6%.

6.1.8.Тексты о политике (на литовском языке)

Всего в указанных текстах было 97847 словоупотреблений. В составленном алфавитно-частотном словнике значилось 16915 разных лексем. В числе наиболее частотных были:

а) из служебного ряда слов: *ir(u)-3169, i(в/на)-1017, kad(что,чтобы)-875, kaip(как)-690, ne(не)-605, iš(из)-577, su(c)-539, apie(про)-376, arba(или)-354,*

tik(только)-350, *už*(за)-301, *jeigu*(если)-293, *bet*(но)-289, *po*(после)-286, *iki*(до)-255, *per*(через)-233, *nuo*(от)-222, *jei*(если)-123, *be*(без)-103 и т.д.;

б) из числа знаменательных частей речи: *rinkimai*(выборы)-2049, *seimas*(сейм)-758, *komisija*(комиссия)-674, *yra*(есть)-622, *tai*(это)-479, *respublika*(республика)-445, *būti*(быть)-389, *galėti*(мочь)-358, *jis*(он)-343, *rinkėjas*(избиратель)-335, *turėti*(уметь)-334, *straipsnis*(статья)-318, *narys*(член)-184, *apylinkė*(округ)-265, *valstybė*(государство)-259, *balsas*(голос)-225, *įstatymas*(закон)-173, *daug*(много)-168, *vėliau*(позже)-166, *darbas*(работа)-156, *prezidentas*(президент)-150 и т.д.;

в) из ряда имён собственных: *Lietuva*-643, *Europa*-68, *Maskva*-60, *AVIA*-58 и т.д.;

По приведенным примерам видно, что высокочастотные слова определяют тематическую сущность данных текстов.

Среди низкочастотных слов отметим такие примеры: *vargstanti*, *vargšas*, *variantas*, *variklis*, *vartoti*, *vartai*, *važiuoti*, *vedėjas*, *veikėjas*, *veiksmingai*, *vėjas*, *vėliava*, *vėluojantis*, *verdiktas*, *versija*, *vertėjas*, *vertybė*, *vertimas*, *vertinimas*, *veržtis*, *vidurinis*, *vidurkis*, *vienetas*, *vienytis*, *vieniša*, *vienkartinė*, *vientisas*, *vietinis*, *vietovė*, *vykimas*, *viltis*, *vyriškis*, *visiškai*, *visuma*, *vojažas*, *žalias*, *žemė*, *žaidimas*, *žemyn*, *žinyba*, *žurnalistai*, *žūti*, *žvilgtelėti* и т.д.

Следовательно, в числе низкочастотных словоупотреблений весьма имеется весьма незначительные признаки тематики изучаемых текстов.

Далее рассмотрим статистическую структуру словника по группам частот (см. табл.№21).

Таблица №21

ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА ТЕКСТОВ ПО ПОЛИТИКЕ

Ранги	Частотная группа	Частота в словнике	Относит. частота	Количество словоупотр.	Относит. частота
1	1 – 4	13908	82,2229	24283	24,8173
2	5 – 9	1562	9,2344	10133	10,3560
3	10 – 100	1343	7,9397	31847	32,5478
4	101 – 1000	99	0,5853	25349	25,9068
5	>1000	3	0,0177	6235	6,3722

Можно утверждать, что и данная структура является «традиционной»: наибольшее число составляет низкочастотная лексика (свыше 82%), наименьшее – высокочастотная (около 8,5%).

6.1.9. Тексты о банковском деле (на литовском языке)

Всего в указанных текстах было 244741 словоупотребления. В составленном алфавитно-частотном словнике значилось 10128 разных лексем. В числе наиболее частотных были:

а) из служебного ряда слов: *ir(u)*-7863, *kad(что,чтобы)*-1663, *iš(из)*-1553, *į(в/на)*-1447, *su(c)*-1429, *nuo(от)*-889, *už(за)*-757, *iki(до)*-737, *kai(когда)*-674, *arba(или)*-546, *per(через)*-483, *bet(но)*-367, *prie(y)*-324, *nors(хотя)*-336 и т.д.;

б) из числа знаменательных частей речи: *bankas(банк)*-3648, *yra(есть)*-2690, *būti(быть)*-2568, *pinigai(деньги)*-1845, *užsienis(заграница)*-1047, *rinka(рынок)*-991, *tai(это)*-973, *jie(они)*-970, *valiuta(валюта)*-774, *politika(политика)*-768, *turėti(иметь)*-691, *galėti(мочь)*-668, *metai(год)*-615, *kapitalas(капитал)*-548, *ekonomika(экономика)*-509, *centrinis(центральный)*-495, *sistema(система)*-465, *kursas(курс)*-459, *šalis(страна)*-441 и т.д.;

в) из ряда имён собственных: *Lietuva*-1913, *Europa*-75), *Euro*-384, *LŽŪB*-237, *Bundesbankas*-192, *Vokietija*-175, *LTB*-99 и т.д.;

Таким образом, высокочастотные слова весьма четко определяют тематическую сущность данных текстов.

Среди низкочастотных слов отметим такие примеры: *straipsnis, studijos, suderinti, sugebėjimas, suklastoti, sunaikinti, suprasti, supirkimas, susieti, suteikti, suvaržytas, svarbesnis, šalis, šeimininkas, tarnyba, tarpininkas, tarptautinis, taupomoji, teikiamas, teisingumas, tekti, telktis, tučia, tikima, tipas, tolesnis, tvirtas, ugdymas, užmokėti, užbaigimas, vadovybė, valstija, vidinis, vykdomas, vystant, žadėdavo, žiaurus, žinia, žlugimas, žinyba, žmonija, žodis* и т.д.

Следовательно, по рядам низкочастотных словоупотреблений тематические признаки лексики изучаемых текстов определить значительно труднее.

Далее рассмотрим статистическую структуру словника по группам частот (см. табл. №22).

Таблица №22

**ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА ТЕМАТИЧЕСКИХ ТЕКСТОВ
ПО БАНКОВСКОМУ ДЕЛУ**

Ранги	Частотная группа	Частота в словнике	Относит. частота	Колич. словоуп.	Относит. частота
1	1 – 4	2997	29,5912	9626	3,9331
2	5 – 9	3536	34,9131	24720	10,1005
3	10 – 100	3258	32,1682	89748	36,6706

4	101 – 1000	321	3,1694	85743	35,0342
5	>1000	16	0,1580	34904	14,2616

Самый примечательный признак данного словника, как и в случае со словником текстов по автомобилизму, является тот факт, что низкочастотная группа словоупотреблений составляет наименьшую часть словника – лишь около 30%. Между тем наибольшую долю словника здесь составляют высокочастотные словоупотребления – свыше 35%. И в данном случае такой результат получен вследствие интенсивного употребления целого ряда тематических терминов.

6.1.10. Тексты о спорте (на литовском языке)

Всего в указанных текстах было 100379 словоупотреблений. В составленном алфавитно-частотном словнике значилось 15472 разные лексемы. В числе наиболее частотных были:

а) из служебного ряда слов: *ir(u)*-3733, *i(в/на)*-787, *su(c)*-619, *kad(что,чтобы)*-555, *iš(из)*-549, *po(после)*-383, *o(a)*-364, *tik(только)*-349, *nuo(от)*-338, *jei(если)*-320, *apie(об)*-302, *bet(но)*-300, *per(через)*-282, *kai(когда)*-256, *iki(до)*-253, *arba(или)*-247, *tačiau(но)*-237, *už(за)*-173, *ant(на)*-173, *be(без)*-149 и т.д.;

б) из числа знаменательных частей речи: *sportas(спорт)*-1009, *yra(есть)*-594, *varžybos(соревнования)*-374, *būti(быть)*-315, *kūnas(тело)*-292, *turėti(иметь)*-285, *galėti(мочь)*-268, *daug(много)*-237, *komanda(команда)*-228, *jie(они)*-220, *kultūra(культура)*-212, *metas(пора)*-196, *labai(очень)*-186, *reikia(надо)*-168, *respublika(республика)*-161, *raumenys(мышцы)*-156, *galima(можно)*-151, *taškas(очко)*-145 и т.д.;

в) имена собственные: *Lietuva*-583, *Europa*-127, *Vilnius*-103, *Kaunas*-79 и т.д.

Таким образом, и в данном словнике высокочастотные слова определяют тематическую сущность текстов.

Среди низкочастотных слов отметим такие примеры: *tartis*, *taryba*, *tarnauti*, *tautinis*, *teikimas*, *taurė*, *technikas*, *tauta*, *teisininkas*, *terpė*, *tezė*, *tikslingas*, *tinklas*, *traktuoti*, *trasa*, *trenerė*, *trumpinimas*, *tuometinė*, *turistinė*, *turistas*, *ugdymas*, *universalus*, *ūkis*, *užkvadratis*, *užregistruotas*, *vaikas*, *valdžia*, *valstybė*, *varyti*, *vartojimas*, *veidas*, *vedėja*, *vėliavėlė*, *veteranas*, *vidurinė*, *vyresnis*, *viršus*, *vora*, *žymiausias*, *žinios*, *žinovai*, *žvaigždė* и т.д.

Следовательно, в числе низкочастотных словоупотреблений можно также найти немало слов, являющихся тематическим признаком изучаемых текстов.

Далее рассмотрим статистическую структуру словника по группам частот (см. табл.№23).

Таблица №23

ЧАСТОТНАЯ СТРУКТУРА СЛОВНИКА ПО ТЕКСТАМ О СПОРТЕ

Ранги	Частотная группа	Частота в словнике	Относит. частота	Количество словоупотр.	Относит. частота
1	1 – 4	11731	75,820	25118	25,0232
2	5 – 9	2038	13,1722	13349	13,2986
3	10 – 100	1612	10,4188	36092	35,9557
4	101 – 1000	89	0,5752	21078	20,9984
5	>1000	2	0,0129	4742	4,7241

Видно, что словник текстов о спорте получился почти традиционным: наибольшая его доля падает на низкочастотные словоупотребления – почти 76%. Высокочастотные словоупотребления составляют около 11% словника.

6.1.11. Параметры и лексический состав сводного словаря

В данном исследовании под словом *словник* понимается список, или перечень слов, полученный в результате лексикостатистической обработки исследуемых тематических текстов. В лексикографических трудах учёных сформировалось понятие *словник* как более узкое по значению в отношении к понятию *словарь* (большая совокупность слов).

Итак, в результате суммирования десяти тематических словников мы получили большой словарь (или совокупность всех тематических словников), который характеризуется такими параметрическими данными:

- а) длинна, или объём, словаря – 98761 разное слово и словоформа;
- б) количество словоупотреблений в мнимых текстах (их суммарной совокупности) – 1306998.

Следовательно, здесь индекс итерации (средняя частота употребления слов в текстах) будет:

$$1306998 : 98761 = 13,23$$

Значит, через каждые 13–14 слов появляется новое слово. По полученному результату можно утверждать, что в весьма большой выборке текстов на родном

(литовском) языке средняя частота словоупотреблений весьма близка к аналогичному показателю, полученному по текстам на русском языке (величина индекса итерации там была равна 11,25). Возможно, это обусловлено тем, что в первом случае выборка текстов на русском языке была чуть меньше (на 94608 словоупотреблений, что составляет по отношению к текстам русского языка 7,8%; а по отношению к выборке текстов на литовском языке – 7,24%.

Полная структура частот отражается в таблице №24. (см.).

Таблица №24

**ЧАСТОТНАЯ СТРУКТУРА СОВОКУПНОГО СЛОВАРЯ ВСЕХ ТЕКСТОВ
НА ЛИТОВСКОМ ЯЗЫКЕ**

Ранги	Частотная группа	Частота в словнике	Процент	Количество словоупотр.	Процент
1	1 – 4	70829	71,7176	118707	9,0824
2	5 – 9	11040	11,1785	72901	5,5777
3	10 – 100	15237	15,4282	402376	30,7863
4	101 – 1000	1537	1,5563	363141	27,7844
5	>1000	118	0,1195	349873	26,7692

По данным таблицы видно, что всё-таки наибольшую часть словаря составляют низкочастотные лексемы – почти 72%. Это, конечно, чуть меньше, чем по отдельным тематическим словникам. Такая же тенденция наблюдается и в отношении среднечастотной группы слов и словоупотреблений.

Но в группе высокочастотных лексем наблюдается обратное явление – их число несколько увеличилось – до 17%.

Приведём примеры высокочастотных словоупотреблений:

а) по ряду служебных слов: *ir(u)*-44344; *i(v/na)*-10696; *kad(что,чтобы)*-10627; *iš(из)*-8516; *su(c)*-8135; *o(a)*-5769; *niu(от)*-4753; *tik(только)*-4590; *tačiau(но)*-3978; *bet(но)*-3737; *už(за)*-3171; *per(через)*-3015; *iki(до)*-2859; *po(после)*-2827; *arba(или)*-2447; *prie(y)*-2091; *jei(если)*-2011; *be(без)*-1915; *nors(хотя)*-1774; *pagal(но)*-1767; *ant(на)*-1718; *jeigu(если)*-1663 и т.д.;

б) по ряду знаменательных частей речи: *yra(есть)*-9796; *būti(быть)*-8023; *jis(он)*-4641; *bankas(банк)*-3652; *galėti(мочь)*-3513; *turėti(иметь)*-3151; *aš(я)*-2944; *labai(очень)*-2455; *metai(годы)*-2287; *rinkimai(выборы)*-2082; *pinigai(деньги)*-2035; *ji(она)*-2169; *metas(пора)*-1814 и т.д.;

в) по ряду имён собственных: *Lietuva*(Литва)-5965; *Europa*(Европа)-1521; *Dakaras*(Дакар)-316 и т.д.

Как видно, в сводном словаре имена собственные как бы “растворились”, т.е. по отношению к большой (совокупной) выборке они перестают доминировать как тематические признаки текстов.

Однако самыми высокими индексами частот отмечается целый ряд служебных слов, так как они являются высокочастотными элементами любых тематических текстов. Краткий обзор совокупного словаря тематических текстов формирует понимание того, как образуется частотный словарь флективных (русского и литовского) языков – лексической системы в целом; какие её элементы являются высокочастотными (активный словарь), а какие – низкочастотными (пассивный словарь).

Чтобы более наглядно можно было убедиться в правдивости высказанной мысли, приведём целый ряд низкочастотных словоупотреблений: *žemas*(низкий)-1*116; *žengti*(шагать)-1*65; *žentas*(зять)-1*9; *žiauna*(челюсть)-1*8; *žiaurumas*(жестокость)-1*1; *žibalas*(керосин)-1*18; *žiburys*(свет)-1*1031; *židinys*(очаг)-1*27; *žievė*(кора)-1*43; *žygis*(шествое)-1*6; *žilas*(седой)-1*37; *žutė*(знак)-1*64; *žindyti*(кормить)-1*47; *žingsnis*(шаг)-1*229; *župys*(жрец)-1*3; *žinovas*(знаток)-1*19; *žiorplys*(ротозей)-1*1; *žirgas*(конь)-1*139; *žiūrėti*(смотреть)-1*342; *žiurkė*(крыса)-1*34; *žtogiškai*(по-человечески)-1*2; *žodis*(слово)-1*1039; *žudymas*(убийство)-1*27; *žūti*(гибнуть)-1*16; *žuvis*(рыба)-1*100 и т.д.

6.1.12. Анализ высокочастотной лексики в текстах на литовском языке

В целях сопоставления наиболее частотной лексики тематических текстов на литовском языке их совокупность рассматривалась с точки зрения заимствований в пределах 500 наиболее употребительных слов (по убывающим частотам).

По отдельным тематическим текстам наибольшее число заимствований можно отметить в словниках по медицине (15,63%), спорту (15,63%), банковскому делу (16,8%); наименьшими относительными числами обозначаются словники текстов по Ветхому Завету (0,78%) и художественной литературе (0,78%). Немного заимствований отмечается также словнике текстов для женщин (4,69%).

Чтобы вышесказанное было очевидным фактом, приведём целый ряд таблиц, в которых заимствования в литовском языке классифицируются в пределах 500 частотных рангов (по их убыванию) и по отдельным темам.

БИБЛИЯ (найденное число слов – 2)

Интервалы (кластеры)	Имена существительные (нарицательные) (В скобках указывается частота в нашем словнике; потом - частота по ЧСРЯ и ЧСЛЯ)
1 – 100	-
101 – 200 (всего 1)	Фараон-(2)142*5*4 gr<egipt.
201 – 300	-
301 – 400 (всего 1)	Фараон-(1)42*5*4 gr<egipt.
401 – 500	-

ХУДОЖЕСТВЕННАЯ ЛИТЕРАТУРА (найденное число слов – 2)

Интервалы (кластеры)	Имена существительные (нарицательные) (В скобках указывается частота в нашем словнике; потом - частота по ЧСРЯ и ЧСЛЯ)
1 – 100	-
101 – 200	-
201 – 300 (всего 1)	Бухгалтер-(1)34*8*45 vok.
301 – 400 (всего 1)	Доктор-(1)25*139*129 lot.
401 – 500	-

МЕДИЦИНА (найденное число слов – 40)

Интервалы (кластеры)	Имена существительные (нарицательные) (В скобках указывается частота в нашем словнике; потом - частота по ЧСРЯ и ЧСЛЯ)
1 – 100 (всего 2)	Алкоголь-(1)109*7*51 arab.; симптом-(1)85*-*13 gr.
101 – 200 (всего 9)	Медицина-(1)77*21*121 lot.; система-(1)75*289*1475 gr.; алкоголизм-(1)68*7*14 arab.; инфекция-(1)68*3*52 lot.; симптом-(1)62*-*13 gr.; организм-(1)59*70*101 lot.; пациентка-(1)56*1 lot.; алкоголик-(1)55*14*8 arab.; доза-(1)54*2*79 gr.
201 – 300 (всего 8)	Алкоголь-(1)109*7*51 arab.; инфекция-(1)49*3*52 lot.; пациент-(1)46*1*52 lot.; операция-(1)45*94*322 lot.; психика-(1)45*5*15 gr.; диагностика-(1)44*1*12 gr.; информация-(1)41*57*702 lot.; орган-(1)39*134*405 gr.
301 – 400 (всего 13)	Алкоголик-(2)67*14*8 arab.; доза-(1)54*2*79 gr.; форма-(1)35*456*956 lot.; информация-(1)35*57*702 lot.; функция-(1)34*27*416 lot.; гормон-(1)34*2*4 gr.; депрессия-(1)33*1*17 lot.; пациент-(1)32*1*52 lot.; паника-(1)32*22*11 gr.; терапия-(1)32*1*7 gr.; алкоголизм-(1)31*7*14 arab.; университет-(1)31*81*386 vok.<lot.; организм-(1)30*70*101 lot.
401 – 500 (всего 8)	Форма-(1)29*456*956 lot.; группа-(1)29*244*934 vok.; концентрация-(1)29*22*133 angl.<lot.; метод-(1)29*170*498 gr.; препарат-(1)29*6*94 lot.; реакция-(1)28*66*171 lot.; результат-(1)28*216*617 pranc.<lot.; диагностика-(1)27*1*12 gr.

АВТОМОБИЛИЗМ (найденное число слов – 35)

Интервалы (кластеры)	Имена существительные (нарицательные) (В скобках указывается частота в нашем словнике; потом - частота по ЧСРЯ и ЧСЛЯ)
1 – 100 (всего 3)	Автомобиль-(6)3920*279*5 gr.+lot.; транспорт-(1)646*69*264 lot.; система-(1)281*289*1475 gr.
101 – 200 (всего 11)	Категория-(1)227*37*84 gr.; класс-(1)222*423*308 lot.; мотоцикл-(2)425*8*31 lot.+gr.; конкурс-(1)214*24*104 lot.; цилиндр-(1)206*28*13 gr.; автомобиль-(1)182*279*5 gr.+lot.; команда-(1)180*180*141 pranc.<lot.; комиссия-(1)172*110*804 lot.; журнал-(1)168*122*168 pranc.; километр-(1)166*251*77 gr.; модель-(1)160*44*248 pranc.<it.<lot.
201 – 300 (всего 6)	Приз-(1)156*10*7 pranc.; конкурс-(1)145*24*104 lot.; купе-(1)127*8*8 pranc.; форма-(1)121*456*956 lot.; формула-(1)117*22*108 lot.; модель-(1)117*44*248 pranc.<it.<lot.
301 – 400 (всего 11)	Команда-(1)112*180*141 pranc.<lot.; момент-(1)108*181*141 lot.; форма-(1)104*456*956 lot.; мотоцикл-(1)104*8*31 lot.+gr.; автомобиль-(2)191*93*5 gr.+lot.; компания-(1)103*73*126 pranc.; конструкция-(1)102*45*173 lot.; система-(1)99*289*1475 gr.; салон-(1)94*25*19 pranc.; чемпионат-(1)91*31*26 angl.; журналист-(1)91*82*157 pranc.
401 – 500 (всего 4)	Автомобиль-(1)82*93*5 gr.+lot.; проект-(1)83*76*713 lot.; контроль-(1)81*80*383 pranc.; модель-(1)117*44*248 pranc.<it.<lot.

ВОЙНА (найденное число слов – 32)

Интервалы (кластеры)	Имена существительные (нарицательные) (В скобках указывается частота в нашем словнике; потом - частота по ЧСРЯ и ЧСЛЯ)
1 – 100 (всего 5)	Республика-(1)298*244*1754 lot.; система-(1)169*289*1475 gr.; министр-(1)147*180*455 lot.; комиссия-(1)140*110*804 lot.; резерв-(1)118*34*49 pranc.<lot.
101 – 200 (всего 8)	Министр-(1)96*180*455 lot.; организация-(1)73*372*1021 pranc.<gr.; политика-(1)68*242*463 gr.; медицина-(1)66*21*121 lot.; геноцид-(1)65*-*14 gr.; администрирование-(1)62*2*22 lot.; министерство-(1)62*133*519 lot.; комиссия-(1)59*110*804 lot.
201 – 300 (всего 10)	Президиум-(1)58*62*92 rus.<lot.; территория-(1)54*95*292 lot.; институция-(2)99*-*262 lot.; экспертиза-(1)49*1*62 lot.; политика-(1)45*242*463 gr.; президент-(1)45*100*395 lot.; акт-(1)43*62*498 lot.; система-(1)43*289*1475 gr.; академия-(1)41*121*294 gr.; функция-(1)41*27*416 lot.
301 – 400 (всего 8)	Министерство-(2)76*133*519 lot.; президент-(1)36*100*395 lot.; план-(1)35*407*358 lot.; пункт-(1)35*80*350 lot.; группа-(1)34*244*934 vok.; нейтралитет-(1)34*-*5 vok.<lot.; бюджет-(1)32*18*278 angl.<pranc.; организация-(1)31*372*1021 pranc.<gr.
401 – 500 (всего 1)	Документ-(1)29*103*1872 lot.

О ЖЕНЩИНАХ (найденное число слов – 12)

Интервалы (кластеры)	Имена существительные (нарицательные) (В скобках указывается частота в нашем словнике; потом - частота по ЧСРЯ и ЧСЛЯ)
1 – 100 (всего 1)	Секс-(3)341*-*5 lot.
101 – 200	-
201 – 300 (всего 4)	Медицина-(1)45*21*121 lot.; порнография-(2)85*-*1 gr.; никотин-(1)42*-*1 pranc.; республика-(1)40*244*1754 lot.
301 – 400 (всего 5)	Акт-(1)35*62*498 lot.; программа-(2)67*179*549 gr.; аборт-(2)68*-*19 lot.; проблема-(1)32*257*726 gr.; сигарета-(1)30*-*59 pranc.
401 – 500 (всего 2)	Проблема-(2)58*257*726 gr.; форма-(1)30*456*956 lot.

СЕЙМ (найденное число слов – 21)

Интервалы (кластеры)	Имена существительные (нарицательные) (В скобках указывается частота в нашем словнике; потом - частота по ЧСРЯ и ЧСЛЯ)
1 – 100 (всего 3)	Коллега-(1)743*29*115 lot.; проект-(1)292*76*713 lot.; комитет-(1)220*160*441 pranc.
101 – 200 (всего 5)	Республика-(1)169*244*1754 lot.; комитет-(1)149*160*441 pranc.; проект-(1)129*76*713 lot.; коллега-(1)120*29*115 lot.; бюджет-(1)105*18*278 angl.<pranc.
201 – 300 (всего 9)	Проблема-(2)143*257*276 gr.; конвенция-(1)79*-*1 lot.; финансы-(1)77*14*137 pranc.<lot.; культура-(1)71*269*946 lot.; ратификация-(1)70*-*5 lot.; комитет-(1)64*160*441 pranc.; департамент-(1)63*12*159 pranc.; комиссия-(1)62*110*804 lot.; кодекс-(1)59*15*161 lot.
301 – 400 (всего 3)	Министерство-(2)106*133*519 lot.; экономика-(1)47*135*311 gr.; проект-(1)42*76*713 lot.
401 – 500 (всего 6)	Транспорт-(1)40*69*265 lot.; вице-министр-(1)40*-*4 lot.; фракция-(1)39*7*121 pranc.<lot.; редакция-(1)39*90*133 pranc.<lot.; университет-(1)39*81*386 vok.; полиция-(1)38*59*465 vok.<gr.

ПОЛИТИКА (найденное число слов – 25)

Интервалы (кластеры)	Имена существительные (нарицательные) (В скобках указывается частота в нашем словнике; потом - частота по ЧСРЯ и ЧСЛЯ)
1 – 100 (всего 4)	Комиссия-(3)1166*110*804 lot.; республика-(1)445*244*1754 lot.; кандидат-(2)312*41*96 lot.; президент-(2)284*100*395 lot.
101 – 200 (всего 6)	Бюллетень-(2)143*7*78 pranc.; комиссия-(1)80*110*804 lot.; документ-(2)125*103*1872 lot.; кандидат-(1)64*41*96 lot.; территория-(1)60*95*292 lot.; конституция-(1)56*27*225 lot.
201 – 300 (всего 5)	Партия-(1)53*532*378 lot.; информирование-(1)52*-*3 lot.; комиссия-(2)93*110*804 lot.; результат-(1)43*216*617 pranc.<lot.; президент-(1)42*100*395 lot.
301 – 400 (всего 8)	Мандат-(1)35*5*27 lot.; кандидат-(1)33*41*96 lot.; министр-(1)33*180*455 lot.; политика-(1)33*242*463 gr.; бюджет-(1)32*18*278 angl.<pranc.; группа-(1)32*244*934 vok.; президент-(1)32*100*395 lot.; протокол-(1)32*10*267 gr.
401 – 500 (всего 2)	Кандидат-(1)28*41*96 lot.; кандидатура-(1)28*41*32 lot.

БАНКОВСКОЕ ДЕЛО (найденное число слов – 43)

Интервалы (кластеры)	Имена существительные (нарицательные) (В скобках указывается частота в нашем словнике; потом - частота по ЧСРЯ и ЧСЛЯ)
1 – 100 (всего 9)	Банк-(8)10140*39*950 it.; политика-(1)768*242*463 gr.; финансы-(1)571*14*137 pranc.<lot.; капитал-(1)548*128*306 angl.<lot.; экономика-(1)509*135*311 gr.; система-(1)465*289*1475 gr.; курс-(1)459*116*138 lot.; кредит-(1)392*6*273 lot.; сектор-(1)361*9*38 lot.
101 – 200 (всего 14)	Кризис-(1)307*38*98 gr.; банк-(2)582*39*950 it.; система-(3)642*289*1475 gr.; резерв-(1)267*34*49 pranc.<lot.; сектор-(1)361*9*38 lot.; сумма-(1)217*62*387 lot.; кредит-(1)207*6*273 lot.; институция-(1)202*-*262 lot.; норма-(1)187*63*338 lot.; политика-(1)185*242*463 gr.; ипотека-(1)180*-* gr.; проблема-(1)180*257*726 gr.; финансы-(1)173*14*137 pranc.<lot.; реструктуризация-(1)173*-*1 lot.
201 – 300 (всего 10)	Банк-(2)326*39*950 it.; капитал-(2)324*128*306 angl.<lot.; институция-(1)164*-*262 lot.; проблема-(2)266*257*726 gr.; норма-(1)133*63*338 lot.; резерв-(1)129*34*49 pranc.<lot.; сектор-(1)121*9*38 lot.; политика-(1)120*242*463 gr.; информация-(1)109*57*702 lot.; сумма-(1)115*62*387 lot.
301 – 400 (всего 7)	Норма-(2)204*63*338 lot.; инвестиция-(1)111*2*85 vok.<lot.; информация-(1)109*57*702 lot.; операция-(2)195*94*322 lot.; республика-(1)105*244*1754 lot.; экономика-(2)191*135*311 gr.; эмиссия-(1)90*1*25 lot.
401 – 500 (всего 3)	Фонд-(1)84*36*3 pranc.<lot.; институция-(1)84*-*262 lot.; проблема-(1)84*257*726 gr.

СПОРТ (найденное число слов – 40)

Интервалы (кластеры)	Имена существительные (нарицательные) (В скобках указывается частота в нашем словнике; потом - частота по ЧСРЯ и ЧСЛЯ)
1 – 100 (всего 4)	Спорт-(1)1009*73*137 angl.; команда-(2)370*180*141 pranc.<lot.; культура-(1)212*269*946 lot.; республика-(1)161*244*1754 lot.
101 – 200 (всего 9)	Центр-(1)84*161*462 lot.; команда-(1)84*180*141 pranc.<lot.; энергия-(1)78*351*183 gr.; калория-(1)72*-* lot.; футбол-(1)66*28*26 angl.; программа-(1)65*179*549 gr.; спортсмен-(1)56*42*38 angl.; спорт-(1)55*73*137 angl.; инвентарь-(1)54*10*27 lot.
201 – 300 (всего 13)	Результат-(1)49*216*617 lot.; метр-(1)47*171*70 pranc.<gr.; группа-(1)46*244*934 vok.; гигиена-(1)46*58*30 gr.; спорт-(1)45*73*137 angl.; аэробика-(1)44*-* angl.; программа-(1)44*179*549 gr.; чемпионат-(1)42*31*43 angl.; клуб-(2)81*121*74 angl.; система-(1)40*289*1475 gr.; автомобиль-(1)39*279*5 gr.+lot.; комитет-(1)38*160*441 pranc.; министр-(1)38*180*455 lot.
301 – 400 (всего 9)	Команда-(1)84*180*141 pranc.<lot.; транспорт-(1)37*69*265 lot.; чемпионат-(1)36*31*43 angl.; код-(1)36*1*176 pranc.<lot.; результат-(1)36*216*617 lot.; атлетика-(1)35*-*3 gr.; спорт-(1)34*73*137 angl.; организация-(1)33*372*1021 pranc.<gr.; гимнастика-(1)32*9*2 gr.
401 – 500 (всего 5)	Спортсмен-(1)31*42*137 angl.; центр-(1)30*161*462 lot.; чемпионат-(1)29*31*43 angl.; система-(1)29*289*1475 gr.; федерация-(1)28*11*63 lot.

Содержание таблиц подтверждает предварительный тезис о том, в каких тематических текстах заимствований было больше всего, а в каких – меньше всего.

Чтобы представить статистическую совокупность заимствований в целом (и по отдельным языкам, откуда эти заимствования появились в литовском языке), рассмотрим сводную таблицу заимствований (см. табл. №25)

Таблица №25

**ЗАИМСТВОВАНИЯ В ТЕКСТАХ НА ЛИТОВСКОМ ЯЗЫКЕ
(суммарная характеристика)**

Интервалы (кластеры)	Лат.	Греч.	Англ.	Немец.	Франц.	Другие языки	Всего
1 – 100	16 13,45%	7 12,07%	2 10,0%	0	4 10,25%	2 20,0%	31 12,15%
101 – 200	30 25,21%	18 31,04%	4 20,0%	0	9 23,08%	3 30,0%	64 25,10%
201 – 300	29 24,37%	14 24,13%	5 25,0%	2 22,22%	11 28,20%	3 30,0%	64 25,10%
301 – 400	31 26,05%	13 22,41%	6 30,0%	4 44,44%	9 23,08%	2 20,0%	65 25,50%
401 – 500	13 10,92%	6 10,35%	3 15,0%	3 33,33%	6 15/38%	0	31 12,15%
Суммы по интервалам	119	58	20	9	39	10	255
Процент по интерва- лам	46,66%	22,75%	7,85%	3,53%	15,29%	3,92%	100%

Как видно, в рассмотренных интервалах частот наибольшее относительное число заимствований отмечается во второй, третьей и четвёртой сотне по убывающим частотным интервалам.

Если рассматривать заимствования по языкам-источникам, то наибольшим числом обозначены заимствования на латинском языке – почти 47%; а на второй позиции – из греческого языка (почти 23%). Третью позицию занимают заимствования из французского языка – свыше 15%. Заметное число заимствований из английского языка – 8%. Из других языков заимствования составляют относительно небольшую долю.

Если сравнить количество заимствований в русском языке по сравнению с таким же явлением в литовском языке, то следует отметить, что здесь наибольшую долю составляют тоже латинизмы – почти 42%. Далее следуют заимствования из греческого языка – около 24%; немало заимствований из французского языка – свыше 19%.

В пределах пятисот наиболее употребительных слов совсем небольшую долю составляют заимствования из немецкого языка – чуть более 1%. Вместе с тем стоит отметить, что заимствования из английского и других языков составляют приблизительно 7%.

Сопоставительный анализ заимствований в литовском и русском языках особенно наглядно показал, что, например, по темам о политике и медицине наблюдается наибольшее

число заимствований. Однако, в литовском языке намного меньше заимствований в текстах по художественной литературе и Библии.

Возможно, сказались результаты большого труда литуанистов по чистке текстов, т.е. бережному отношению к родному языку и их усилия по реализации ресурсов литовского языка.

Таковы сопоставительные итоги исследования заимствований в пределах пятисот наиболее употребительных слов по убывающим интервалам (десяти выборок разной тематики текстов на литовском и такого же количества – на русском языках).

7. Сопоставительный анализ высокочастотной русско-литовской лексики

В исследованиях частот словоупотреблений заметны сходства и различия в разных языках. Обычно высокочастотные слова находят аналоги и в лексических системах других языков.

В нашем случае по теме “Библия” высокочастотными были такие словоупотребления:

а) из числа имён существительных: *день(diena)* – 207 (из текста источника); 580 (в ЧСЛЯ); 1345 (в ЧСРЯ);

земля(žetė) – (соответственно) 181-474-1074; *человек(žmogus)* – 135-1421-1677.

б) из числа глаголов: *быть(būti)* – 900-4167-13307; *иметь(turėti)* – 269-988-846; *говорить(kalbėti)* – 146-335-2573.

По теме “Художественная литература” высокими частотами были отмечены во всех трёх сопоставительных источниках такие словоупотребления: *человек(žmogus)*; *рука(ranka)*; *говорить(kalbėti)*; *смотреть(žiūrėti)*; *хотеть(norėti)* и т.д.;

из числа наречий: *сейчас(dabar)*; *здесь(čia)*; *очень(labai)*; *хорошо(gerai)* и т.д.

Полная картина примеров приводится в виде иллюстративных тематических таблиц, представленных в приложениях.(см. табл.)

Таким образом, высокочастотные словоупотребления в русском языке находят такие же соответствия в лексике литовского языка. Однако, можно найти немало примеров с обратным статистическим значением, когда тематические слова обозначаются высокими индексами, а их соответствия в частотных словарях литовского и русского языков обозначаются весьма низкими индексами:

Заграница (užsienis) – 1047-212-2; *валюта (valiuta)* – 774-26-9; *банк (bankas)* – 368-28-39; *финансы (finansai)* – 571-17-14 и т.д.

Само собой разумеется, что здесь речь идёт о терминах банковского дела, что и определяет их высокую частоту употребления в тематических текстах и относительно низкую частоту в частотных словарях.

Следовательно, на основе приведённых примеров мы ещё раз убедились в том, что частота словоупотреблений зависит от объёма выборки и её тематики.

8. Сопоставительный анализ низкочастотной русско-литовской лексики

Как в текстах на русском языке, так и в текстах на литовском языке чётко обозначается структура словника, в которой наибольшая доля падает на низкочастотные словоупотребления. По нашим подсчётам низкочастотная лексика с индексами от 1 до 4 включительно в текстах на литовском языке составила в среднем 82,33%; в текстах на русском языке – 78,54%. Следовательно, разность – всего лишь в пределах 3,79%.

Рассмотрим ряды низкочастотной лексики:

а) в случаях, когда словоупотребления обозначаются низкими индексами частот в текстах на сопоставляемых языках и в соответствующих частотных словарях указанных языков (после слова приводится перевод и частота в текстах на русском языке, в ЧСРЯ и ЧСЛЯ): *эстет* – 1-1-3; *этикет* – 1-1-6; *этиология* – 1-0-2; *этнография* – 1-3-6; *эфирный* – 1-2-1; *ювелир* – 1-3-2; *ювелирный* – 1-0-1; *юмореска* – 1-1-4; *юмористический* – 1-4-2; *юрта* – 1-1-2; *ящер* – 1-2-1 и т.д.;

б) в случаях, когда в тематических текстах на русском языке словоупотребление является низкочастотным, а в ЧСРЯ и ЧСЛЯ – высокочастотным: *этаж* – 2-40-91; *этап* – 1-57-196; *этикетка* – 1-11-15; *эффективность* – 1-32-120; *эффект* – 1-37-90; *юбилей* – 1-26-19; *юный* – 3-44-510; *юность* – 1-72-71; *юноша* – 2-65-93; *яблоня* – 2-20-40; *ягода* – 1-11-30; *ядерный* – 2-84-54; *ядовитый* – 2-16-15; *ядро* – 1-51-44; *яйцо* – 1-25-65; *ящик* – 1-78-64 и т.д.

Таким образом, видно, что не все низкочастотные словоупотребления являются такими же в лексических системах сопоставляемых языков. Всё же можно утверждать, что немалый ряд низкочастотных словоупотреблений в исследуемых текстах были такими же и в лексических системах русского и литовского языков, о чём свидетельствуют низкие индексы частот как в ЧСРЯ, так и в ЧСЛЯ.

9. Обобщение результатов исследования и выводы

Сопоставительный анализ лексики текстов разной тематики методами лингвистической статистики даёт большое количество разнообразных результатов в отношении функционирования слов в русском и родном (литовском) языках. В процессе исследования разнообразных текстов выяснилось, что лексическая

система русского языка является более открытой для заимствований. Такой тезис не совсем был бы объективным по отношению к литовскому языку, так как в ряде текстов различной тематики было обнаружено заметно меньше заимствований. Возможно, сказалась интенсивная работа литуанистов, которые формируют общественное мнение, что к родному языку надо относиться бережно, т.е. в первую очередь пользоваться лексическими ресурсами своего языка.

При составлении тематических словников, всякий раз вырисовывается их статистическая структура:

а) наибольшую долю словника составляют низкочастотные слова;
б) наименьшую долю словника составляет высокочастотная лексика;
в) тематические термины и связанная с ними лексика в основном занимают высокие ранговые позиции в любом словнике текстов;

г) в разных словниках тематических текстов самые высокие частотные позиции занимают наиболее интенсивно употребляемые союзы, предлоги, некоторые местоимения и наречия, составляя остриё, или верхушку, «статистической пирамиды» словника;

д) характерным признаком узкоспециальных текстов (например, по банковскому делу, автомобилизму) является тот факт, что при наличии одинаковых параметров их словники формируются намного меньше по объёму, чем в обычных случаях; Такое происходит за счёт того, что чрезвычайно часто, интенсивно, употребляется специализированная лексика, т.е. термины данной области. Поэтому в таких словниках низкочастотная лексика составляет меньшую долю, чем в обычных лексикостатистических исследованиях.

Сопоставительное исследование функционирования лексики в русском и литовском языках имеет общие и различающиеся признаки. Иначе говоря, на основе сотысячных выборок нельзя прямолинейно утверждать, что однозначные слова обязательно будут отмечены схожими индексами частот, которые попадут в одни и те же частотные группы. При этом следует подчеркнуть, что признаками схожести всегда отличается высокочастотная лексика. Наоборот, весьма значительными признаками различий отмечается низкочастотная лексика.

Тематика интернетных текстов была весьма случайной. Она никак не покрывает те принципы, которые кладутся в основу составления частотных словарей, где учитываются функционально-речевые сферы (например,

художественная проза, драматургия, научные тексты, публицистические тексты). Поэтому совокупность текстов по десяти темам, составляя не менее одного миллиона словоупотреблений, не является тождественным результатом по отношению к ЧСРЯ. Конечно, в указанном источнике сверки учитывались исходные формы слов, а в нашем исследовании процесс лемматизации (превращения, трансформации словоформ в исходные, или словарные, формы) не производился в виду того, что это очень большая по объёму работа. Вот почему в нашем сводном частотном словаре литовского языка по десяти темам значится 98761 лексема, а в “Частотном словаре литовского языка” – 38337 разных слов, т.е. словник источника сопоставления в 2,6 раза меньше по длине, хотя выборка в целом, можно сказать, была одинаковой: 1200000 словоупотреблений в текстах на литовском языке; 1306998 – в нашем исследовании.

Между тем выборка текстов на русском языке составила 1212390, на основе которой был получен словарь объёмом в 107775 разных лексем. Таким образом, словник текстов на русском языке в нашем исследовании превысил аналог на литовском языке на 9014 лексем, что составило 9,1%.

В отношении величины выборки результаты были иными: объём текстов на литовском языке, или выборка, была на 94608 словоупотреблений больше, что составляет 7,8% по отношению к объёму текстов на русском языке.

Значит, относительно равные выборки не дают повода ожидать, что будут такими же равными сформированные их словники. Предположительно можно сказать, что такое явление зависит от грамматической структуры сравниваемых языков. Например, в литовском языке глаголы 3-его лица единственного и множественного чисел совпадают (*jis rašo; jie rašo*), поэтому глагол “*rašo*” фиксируется компьютерной программой как тождественная, одинаковая единица счёта с суммарным индексом частоты. В русском же языке данный глагол по 3-ему лицу даёт две разные словоформы (*он пишет; они пишут*). Поэтому компьютерная программа считает их как две разные лексемы.

В процессе исследования познаются разные тонкости, которые можно получить на основе методов лингвистической статистики. Нами были определены различия и общие характеристики выборок текстов на русском и литовском языках. В каждом случае определяется своеобразие сопоставляемых языков.

Можно предполагать, что знакомство с основными результатами исследования может быть полезным в прикладных целях, например, в лингводидактике, при отборе лексического материала для учебных целей на основе дифференциации словоупотреблений по величине индексов частот.

Bibliografija

1. Алексеев П.Н. Статистическая лексикография. Л., ЛГПИ, 1975.
2. Библийская энциклопедия. Репринтное издание., М., 1991.
3. Большой толковый словарь русского языка. «Норинт». – 1998.
4. Головин Б.Н. Язык и статистика. «Просвещение», 1971.
5. Лемхенас Х. Русско-литовский словарь. Т.1-4. – В. «Мокслас». – 1982.
6. Либерис А. Литовско-русский словарь. В., «Мокслас». – 1988.
7. Лингвистический энциклопедический словарь. М., 1990.
8. Ожегов С.И. Словарь русского языка, М., 1978
9. Орфографический словарь русского языка, М., 1996.
10. Пиотровский Р.Г. Текст. Машина. Человек. М., 1975.
11. Пиотровский Р.Г., Бектаев К., Пиотровская А. Математическая лингвистика, М., 1977.
12. Русский язык. Энциклопедия. М., 1979.
13. Словарь иностранных слов. М., 1985.
14. Словарь языка Пушкина. Т.1-4. М., 1956-1961.
15. Частотный словарь русского языка. (ЧСРЯ) под ред.Л.Засориной. М., 1977.
16. Частотный словарь современного русского языка (ЧССРЯ).Под ред.Л.Леннгрена.1993.
17. Grumadienė L., Žilinskienė V. Dažninis dabartinės rašomosios lietuvių kalbos žodynas. – Mokslo ir enciklopedijų leidybos institutas. – V. – 1998.
18. Korsakas J. Definicijų leksika*Лексика дефиниций*Lexicon of Definitions. Šiauliai.-1998.
19. Lietuvių kalbos enciklopedija. V., 1999.
20. Tarptautinių žodžių žodynas. V.,1985.
21. Žilinskienė V. Lietuvių kalbos dažninis žodynas. V., Mokslas. – 1990.

INTERNETO TEKSTŲ LIETUVIŲ IR RUSŲ KALBOMIS LEKSIKOS GREVINAMASIS LINGVOSTATISTINIS TYRIMAS

Santrauka

Aptariami rusų ir lietuvių kalba paskelbtų tekstų leksikos lingvostatistinės analizės rezultatai. Abiem kalbomis buvo išnagrinėta po 10 įvairios tematikos tekstų, kurių imtys – po 100000 žodžių pavartojimų. Sudarius kiekvieno teksto teminius žodynėlius, norėta sužinoti, kokia bus visumos žodyno apimtis ir sudėtis, koks bus jų santykis su išleistais lietuvių bei rusų kalbų dažniniais žodynais (didelio ir mažo žodžių dažnumo atžvilgiu).

Sudarytų teminių žodynėlių apimtis – vidutiniškai daugiau kaip 19000 skirtingų žodžių ir jų formų. Pažymėtina, jog rusų kalba parašytų tekstų žodynėliuose skirtingų žodžių yra maždaug 110 daugiau. Galima manyti, kad toks skirtumas atsiranda dėl gramatinių kategorijų raiškos skirtumų (pavyzdžiui, lietuvių kalbos vienaskaitos ir daugiskaitos 3-io asmens formos sutampa, taigi skirtingų leksemų skaičius mažėja).

Tarptautinių žodžių grevinamoji analizė parodė, jog jų esama daugiau rusų kalbos teminiuose tekstuose. Padaryta išvada, kad lituanistikos atstovų pastangos keisti kai kuriuos tarptautinius žodžius gimtosios kalbos leksinės sistemos atitikmenimis yra rezultatyvus kalbos kultūros puoselėjimi būdas.

Teminių tekstų žodynėliuose didžiausiais dažnumų indeksais žymimi terminai ir su jais susiję žodžiai.

Įsitikinta, kad yra prasminga lingvistinės statistikos metodais (pagal dažnumų dydžius) skirstyti leksiką į aktyvųjį ir pasyvųjį žodyną. Tai siejama su lingvodidaktikos uždaviniais.

1. Biblija (Senasis Testamentas)

№	Žodis lietuvių kalba	Žodžio vertimas į rusų kalbą	Žod. dažnis šaltinyje	Žod. dažnis LKDŽ	Žod. dažnis RKDŽ
	(10 daiktav.)	(10 daiktav.)			
1	Viešpats	Господь	881	2	137
2	Dievas	Бог	515	23	474
3	Diena	День	207	580	1345
4	Žemė	Земля	181	474	1074
5	Kraštas	Край	174	102	187
6	Sūnus	Сын	168	67	332
7	Kunigas	Священник	163	6	47
8	Akivaizda	Очевидность	150	13	3
9	Atnaša	Приношение	139	2	-
10	Žmogus	Человек	135	1421	1677
	(10 veiksm.)	(10 veiksm.)			
11	Būti	Быть	900	4167	13307
12	Tarti	Произнести	364	100	49
13	Turėti	Иметь	269	988	846
14	Kalbėti	Говорить	146	335	2573
15	Padaryti	Сделать	131	204	732
16	Sakyti	Сказать	92	325	2909
17	Eiti	Идти	81	175	1818
18	Valgyti	Кушать	69	15	32
19	Paimti	Взять	60	51	328
20	Ateiti	Прийти	56	197	543
	(5 būdvardžiai)	(5 būdvardžiai)			
21	Nešvarus	Нечистый	78	6	13
22	Švarus	Чистый	33	14	225
23	-	-	-	-	-
24	-	-	-	-	-
25	-	-	-	-	-
	(5 prieveiksm.)	(5 prieveiksm.)			
26	Drauge	Вместе	103	35	500
27	Dabar	Сейчас	77	438	1210
28	Daug	Много	61	1111	664
29	Niekada	Никогда	44	56	368
30	Viršum	Сверху	43	36	61

2. Grožinė literatūra

№	Žodis lietuvių kalba	Žodžio vertimas į rusų kalbą	Žod. dažnis šaltinyje	Žod. dažnis LKDŽ	Žod. dažnis RKDŽ
	(10 daiktav.)	(10 daiktav.)			
1	Tėvas	Отец	125	82	605
2	Galva	Голова	120	95	849
3	Žmogus	Человек	119	1421	1677
4	Akis	Глаз	85	123	1093
5	Veidas	Лицо	82	67	745
6	Vyras	Мужчина	78	215	110
7	Ranka	Рука	76	228	1596
8	Moteris	Женщина	73	325	429
9	Motina	Мать	73	117	564
10	Durys	Дверь	71	44	733
	(10 veiksm.)	(10 veiksm.)			
11	Būti	Быть	680	4167	13307
12	Tarti	Произнести	144	100	49
13	Turėti	Иметь	106	988	846
14	Sakyti	Говорить	91	325	2573
15	Žiūrėti	Смотреть	78	142	818
16	Stovėti	Стоять	69	55	922
17	Paklausti	Спросить	67	39	553
18	Norėti	Хотеть	66	330	1295
19	Galėti	Мочь	65	920	3373
20	Atrodyti	Выглядеть	63	178	70
	(5 būdvardžiai)	(5 būdvardžiai)			
21	Šaltas	Холодный	25	38	241
22	-	-	-	-	-
23	-	-	-	-	-
24	-	-	-	-	-
25	-	-	-	-	-
	(5 prieeiksm.)	(5 prieeiksm.)			
26	Dabar	Сейчас	289	438	1210
27	Čia	Здесь	286	655	1161
28	Nieko	Ничего	153	215	462
29	Labai	Очень	109	876	1415
30	Gerai	Хорошо	95	420	832

3.Medicina

№	Žodis lietuvių kalba	Žodžio vertimas į rusų kalbą	Žod. dažnis šaltinyje	Žod. dažnis LKDŽ	Žod. dažnis RKDŽ
	(10 daiktav.)	(10 daiktav.)			
1	Metas	Пора	236	389	509
2	Liga	Болезнь	214	56	92
3	Smegenys	Мозг	192	5	61
4	Gydymas	Лечение	158	19	25
5	Galva	Голова	131	95	849
6	Vaikas	Ребёнок	130	468	284
7	Amžius	Возраст	116	162	100
8	Alkoholis	Алкоголь	109	6	7
9	Vaistas	Лекарство	106	17	27
10	Žmogus	Человек	105	1421	1677
	(10 veiksm.)	(10 veiksm.)			
11	Galėti	Мочь	558	920	3373
12	Būti	Быть	434	4167	13307
13	Turėti	Иметь	181	988	846
14	Gerti	Пить	93	21	243
15	Vartoti	Принимать	89	19	185
16	Padėti	Помочь	68	313	249
17	Nustatyti	Установить	66	98	120
18	Skirti	Отделить	65	307	13
19	Priklausyti	Принадлежать	60	123	94
20	Sirgti	Болезнь	52	22	77
	(5 būdvardžiai)	(5 būdvardžiai)			
21	Kartus	Горкий	68	9	-
22	Įvairus	Разнообразный	42	284	132
23	Sergantis	Больной	29	-	134
24	Ūminis	Внезапный	29	-	43
25	Lytinis	Половой	27	9	5
	(5 prievoksm.)	(5 prievoksm.)			
26	Labai	Очень	239	876	1415
27	Dažnai	Часто	155	274	298
28	Daug	Много	130	1111	664
29	Būtinai	Обязательно	108	25	133
30	Kartu	Вместе	89	-	500

4. Automobiliizmas

№	Žodis lietuvių kalba	Žodžio vertimas į rusų kalbą	Žod. dažnis šaltinyje	Žod. dažnis LKDŽ	Žod. dažnis RKDŽ
	(10 daiktav.)	(10 daiktav.)			
1	Automobilis	Автомобиль	1278	51	93
2	Metai	Год	648	1563	2167
3	Transportas	Транспорт	646	58	69
4	Variklis	Двигатель	633	28	70
5	Lenktynės	Гонка	445	28	30
6	Kartas	Раз	338	756	1230
7	Vairas	Руль	330	14	26
8	Gamintojas	Производитель	297	10	15
9	Kėbulas	Кузов	295	-	10
10	Sistema	Система	281	183	289
	(10 veiksm.)	(10 veiksm.)			
11	Būti	Быть	1220	4167	13307
12	Turėti	Иметь	661	988	846
13	Galėti	Мочь	506	920	3373
14	Atrodyti	Выглядеть	246	178	70
15	Sukurti	Создать	245	186	318
16	Tapti	Стать	212	310	1549
17	Reikėti	Нуждаться	207	658	49
18	Laimėti	Выиграть	162	85	35
19	Pavykti	Удаться	156	49	158
20	Pasiekti	Достать	126	224	127
	(5 būdvardžiai)	(5 būdvardžiai)			
21	Naujas	Новый	231	790	1722
22	Maksimalus	Максимальный	150	7	22
23	Techninis	Технический	130	111	183
24	Benzininis	Бензиновый	87	2	5
25	Didelis	Большой	80	697	2066
	(5 prievoksm.)	(5 prievoksm.)			
26	Beveik	Почти	365	189	507
27	Daug	Много	248	1111	664
28	Visiškai	Совсем	213	140	521
29	Toli	Далеко	192	276	342
30	Labai	Очень	177	876	1415

5. Apie karą

№	Žodis lietuvių kalba	Žodžio vertimas į rusų kalbą	Žodžio dažnis šaltinyje	Žodžio dažnis LKDŽ	Žodžio dažnis RKDŽ
	(10 daiktav.)	(10 daiktav.)			
1	Karas	Война	1147	255	825
2	Apsauga	Охрана	791	48	64
3	Tarnyba	Служба	776	38	167
4	Kraštas	Край	751	102	187
5	Valstybė	Государство	365	273	460
6	Respublika	Республика	298	714	244
7	Straipsnis	Статья	290	113	230
8	Karys	Воин	282	18	30
9	Kariuomenė	Войско	245	53	7
10	Įstatymas	Закон	226	77	315
	(10 veiksmaz.)	(10 veiksmaz.)			
11	Būti	Быть	617	4167	13307
12	Galėti	Мочь	331	920	3373
13	Turėti	Иметь	150	988	846
14	Vykdyti	Выполнять	82	206	136
15	Privalėti	Быть обязанным	70	78	48
16	Atlikti	Исполнить	68	227	21
17	Reikėti	Нуждаться	61	658	49
18	Tvirtinti	Утверждать	46	51	68
19	Gauti	Получить	44	299	446
20	Nustatyti	Установить	43	98	120
	(5 būdvardžiai)	(5 būdvardžiai)			
21	Profesinis	Профессиональный	228	59	32
22	Karinis	Военный	94	117	485
23	Pradinis	Начальный	74	25	24
24	Civilinis	Штатский	64	17	17
25	Mėnesinis	Месячный	50	24	6
	(5 prieveiksm.)	(5 prieveiksm.)			
26	Mažai	Мало	91	181	316
27	Daug	Много	67	1111	664
28	Kartu	Вместе	64	2	500
29	Vėlai	Поздно	55	116	178
30	Ilgai	Долго	44	78	305

6. Straipsniai moterims

№	Žodis lietuvių kalba	Žodžio vertimas į rusų kalbą	Žod. dažnis šaltinyje	Žod. dažnis LKDŽ	Žod. dažnis RKDŽ
	(10 daiktav.)	(10 daiktav.)			
1	Moteris	Женщина	310	325	429
2	Smurtas	Насилие	245	10	46
3	Metai	Год	210	1563	2167
4	Gimda	Матка	159	1	1
5	Dievas	Бог	139	23	474
6	Vaikas	Ребёнок	132	468	284
7	Seksas	Секс	123	2	-
8	Žmogus	Человек	121	1421	1677
9	Gyvenimas	Жизнь	112	548	1547
10	Meilė	Любовь	109	132	140
	(10 veiksmaz.)	(10 veiksmaz.)			
11	Būti	Быть	920	4167	13307
12	Galėti	Мочь	402	920	3373
13	Turėti	Иметь	272	988	846
14	Reikėti	Нуждаться	136	658	49
15	Norėti	Хотеть	130	330	1295
16	Tapti	Стать	82	310	1549
17	Atsirasti	Появиться	71	90	208
18	Sukelti	Вызвать	69	82	162
19	Apsaugoti	Защитить	61	16	17
20	Suteikti	Предоставить	53	116	35
	(5 būdvardžiai)	(5 būdvardžiai)			
21	Lytinis	Половой	86	9	5
22	Seksualinis	Сексуальный	50	3	-
23	Fizinis	Физический	32	58	89
24	Didelis	Большой	31	697	2066
25	-	-	-	-	-
	(5 prieveiksm.)	(5 prieveiksm.)			
26	Labai	Очень	171	876	1415
27	Daug	Много	155	1111	664
28	Dabar	Сейчас	134	438	1210
29	Dažnai	Часто	95	274	298
30	Ypač	Особенно	82	257	375

7. Seimas

№	Žodis lietuvių kalba	Žodžio vertimas į rusų kalbą	Žod. dažnis šaltinyje	Žod. dažnis LKDŽ	Žod. dažnis RKDŽ
	(10 daiktav.)	(10 daiktav.)			
1	Pirmininkas	Председатель	932	208	233
2	Įstatymas	Закон	871	77	315
3	Seimas	Сейм	753	-	-
4	Kolega	Коллега	743	24	29
5	Valstybė	Государство	394	273	460
6	Straipsnis	Статья	388	113	230
7	Projektas	Проект	292	51	76
8	Narys	Член	222	154	194
9	Darbas	Работа	218	1759	1103
10	Vyriausybė	Правительство	206	164	267
	(10 veiksm.)	(10 veiksm.)			
11	Būti	Быть	1738	4167	13307
12	Manyti	Думать	282	115	1094
13	Turėti	Иметь	202	988	846
14	Galėti	Мочь	199	920	3373
15	Norėti	Хотеть	188	330	1295
16	Priimti	Принять	176	131	328
17	Pasakyti	Сказать	157	304	2909
18	Priimti	Припомнить	145	59	328
19	Kviesti	Звать	142	46	190
20	Pritarti	Одобрить	137	40	23
	(5 būdvardžiai)	(5 būdvardžiai)			
21	Valstybinis	Государственный	48	128	240
22	Socialinis	Социальный	46	185	142
23	Aukštas	Высокий	45	404	549
24	-	-	-	-	-
25	-	-	-	-	-
	(5 prieveiksm.)	(5 prieveiksm.)			
26	Dabar	Сейчас	456	438	1210
27	Labai	Очень	432	876	1415
28	Šiandien	Сегодня	189	217	497
29	Daug	Много	155	1111	664
30	Toli	Далеко	137	276	342

8. Politika

№	Žodis lietuvių kalba	Žodžio vertimas į rusų kalbą	Žod. dažnis šaltinyje	Žod. dažnis LKDŽ	Žod. dažnis RKDŽ
	(10 daiktav.)	(10 daiktav.)			
1	Rinkimai	Выборы	2049	-	19
2	Seimas	Сейм	758	-	-
3	Komisija	Комиссия	674	99	110
4	Respublika	Республика	445	714	244
5	Rinkėjas	Выбиратель	335	21	-
6	Straipsnis	Статья	318	113	230
7	Apygarda	Округ	311	16	12
8	Apylinkė	Окрестность	265	39	17
9	Valstybė	Государство	259	273	460
10	Balsas	Голос	225	56	587
	(10 veiksm.)	(10 veiksm.)			
11	Būti	Быть	622	4167	13307
12	Galėti	Мочь	358	920	3373
13	Turėti	Иметь	334	988	846
14	Nustatyti	Установить	98	98	120
15	Pateikti	Предоставить	67	126	35
16	Priimti	Принять	64	131	328
17	Kalbėti	Говорить	52	335	2573
18	Sudaryti	Составить	51	252	104
19	Gauti	Достать	48	299	446
20	Skelbti	Объявлять	42	31	70
	(5 būdvardžiai)	(5 būdvardžiai)			
21	Politinis	Политический	52	209	302
22	Vyriausiasis	Верховный	49	-	66
23	-	-	-	-	-
24	-	-	-	-	-
25	-	-	-	-	-
	(5 prieveiksm.)	(5 prieveiksm.)			
26	Daug	Много	168	1111	664
27	Vėlai	Поздно	166	116	178
28	Mažai	Мало	121	181	316
29	Dabar	Сейчас	98	438	1210
30	Labai	Очень	96	876	1415

9. Bankininkystė

№	Žodis lietuvių kalba	Žodžio vertimas į rusų kalbą	Žod. dažnis šaltinyje	Žod. dažnis LKDŽ	Žod. dažnis RKDŽ
	(10 daiktav.)	(10 daiktav.)			
1	Bankas	Банк	3648	29	39
2	Pinigai	Деньги	1845	47	323
3	Užsienis	Заграница	1047	212	2
4	Rinka	Рынок	991	37	54
5	Paskola	Заём	983	-	-
6	Valiuta	Валюта	774	26	9
7	Palūkanos	Проценты	772	-	172
8	Politika	Политика	768	140	242
9	Metai	Год	615	1563	2167
10	Finansai	Финансы	571	17	14
	(10 veiksm.)	(10 veiksm.)			
11	Būti	Быть	2690	4167	13307
12	Turėti	Иметь	691	988	846
13	Galėti	Мочь	668	920	3373
14	Sudaryti	Составить	201	252	104
15	Reikėti	Нуждаться	158	7	49
16	Palyginti	Сравнить	148	62	43
17	Gauti	Получить	119	299	446
18	Pradėti	Начать	119	319	540
19	Tapti	Стать	109	310	1549
20	Naudoti	Использовать	102	143	180
	(5 būdvardžiai)	(5 būdvardžiai)			
21	Centrinis	Центральный	495	29	137
22	Vertybinis	Достойный	316	6	48
23	Valstybinis	Государственный	312	128	240
24	Komercinis	Коммерческий	219	8	-
25	Privatus	Частный	197	10	60
	(5 prieveiksm.)	(5 prieveiksm.)			
26	Labai	Очень	591	876	1415
27	Daug	Много	363	1111	664
28	Paprastai	Просто	158	71	553
29	Visiškai	Совсем	154	140	521
30	Anksti	Рано	148	140	183

10. Sportas

№	Žodis lietuviškai	Žodžio vertimas į rusų kalbą	Žod. dažnis šaltinyje	Žod. dažnis LKDŽ	Žod. dažnis RKDŽ
	(10 daiktav.)	(10 daiktav.)			
1	Sportas	Спорт	1009	345	73
2	Varžybos	Соревнования	374	259	66
3	Kūnas	Тело	292	46	310
4	Komanda	Команда	228	162	180
5	Kultūra	Культура	212	301	269
6	Metas	Пора	196	389	509
7	Respublika	Республика	161	714	244
8	Raumuo	Мускул	156	4	12
9	Krepšinis	Баскетбол	126	36	1
10	Moksleivis	Учащийся	126	121	18
	(10 veiksm.)	(10 veiksm.)			
11	Būti	Быть	594	4167	13307
12	Turėti	Иметь	285	988	846
13	Galėti	Мочь	268	920	3373
14	Dalyvauti	Участвовать	76	207	92
15	Daryti	Делать	73	194	718
16	Privalėti	Быть обязанным	57	78	48
17	Lankyti	Посещать	55	89	14
18	Vykti	Направляются	53	212	45
19	Naudoti	Использовать	44	143	180
20	Skirti	Разлучить	40	307	2
	(5 būdvardžiai)	(5 būdvardžiai)			
21	Įvairus	Разнообразный	42	284	132
22	Bendras	Общий	41	232	401
23	Lengvas	Лёгкий	32	124	152
24	Didelis	Большой	28	697	2066
25	-	-	-	-	-
	(5 prievieksm.)	(5 prievieksm.)			
26	Daug	Много	237	1111	664
27	Labai	Очень	186	876	1415
28	Mažai	Мало	69	181	316
29	Paprastai	Просто	50	71	553
30	Ypač	Особенно	46	257	375

Интернетные адреса:

Straipsniai moterims

1. www.beuty.lt “Vienos nakties nuotykis”, “Griežtai apie rimtus dalykus”, “Vaiko miegas”;
2. www.venera.lt “Seksas”;
3. www.kvsc.lt “Gimdos kaklelio vėžys”;
4. www.geocities.lt “Kaip išvengti išprievartavimo?”;
5. margiris.w3.lt “N.Murnikovienė ir jos treniruotė”, “Pilvo raumenų treniruotė”;
6. www3.vdu.lt “Meilė ir skaistybė santuokoje”, “Seksas”, “Lytiniai santykiai yra šventi”, “Pornografija”, “Pasėkmės”, “Prostitucija”, “Skaistumas”, “Kova už tyrumą”, “Abortas”, “Dirbtinis apvaisinimas”, “Poabortinis sindromas”, “Pseudoargumentai žudyti”, “Dešimt priešasčių turėti dar vieną vaiką”, “Kontracepcija”;
7. www.vaikas.lt “Moters – motinos gimimas”;
8. www.cosmopolitan.lt “Tik moterims: liga, kelianti pavojų”;
9. sveikatos.tinklapis.lt “7 kiekvienai moteriai svarbūs sveikatos tyrimai”;
10. www.is.lt “Anglai padės jaunimui ir skriaudžiamoms moterims”;
11. www.marie.lt “Moterims”;
12. www.medcentras.lt “Šiuolaikinė moteris privalo rūpintis savo krūtims”;
13. www.lygus.lt “Smurtas prieš moteris”;
14. www.is.lt “Homoseksuali moteris Lietuvoje”;
15. www3.lrs.lt “Smurtas prieš moteris”;
16. www.qedata.lt “Įžeista moteris”;
17. www.policija.lt “Prievartos prieš moteris problematika Lietuvos žiniasklaidoje”;
18. groups.msn.com “Moterų nevaisingumo priežastys”;
19. www.sociumas.lt “Moterų padėtis apmokamo darbo rinkoje”;
20. www.sveikuoliai.lt “Moterų fizinė būklė”;
21. www.moterukoalicija.webinfo.lt “Projektas”;
22. www.mamos_svetaine.de “Moters ciklas”;
23. www2.omnitel.net “Meilė ir seksas pagal Bibliją”.

Статьи для женщин

1. www.ivanogo.ac.ru “Гендерные отношения в русской поэзии 17-18 в.в.”, “О положении женщин в республике Кореллия”, “Материнство как социально исторический феномен”, “Несколько вариантов женской самореализации”, “Одарённость и гендер”, “Технологии гендера рекламных дискурсивных практик”, “Современные гендерные исследования в Дании”;
2. www.4woman.ru “У женщин мозги другие?”, “Женские страхи”;
3. www.eva.ru “Женские претензии в постели”, “Шёпот нижнего белья”, “Хочу богатого мужа”;
4. www.sarafan.ru “Внутри сумки”;
5. www.podrugа.ru “Хочу быть стервой”, “С любимыми не расставайтесь”, “Любовь по интернету”, “Операция “О”, “Казусные ситуации”, “Свет мой, зеркальце, скажи...”, “Весенние страдания”;
6. www.djin.com.ua “Мужчины обязаны дарить бельё женщинам. Почему? 28 причин”;
7. doctor-sex.ru “Бесплодие”, “Мамология”;
8. berk.msk.ru “Венерические заболевания”, “Гинеколог”.

Sportas

1. www.extremesport.lt;
2. www.parasparniai.mums.lt;
3. www.climbing.lt;
4. www.margins.w3.lt;
5. www1.omnitel.net;
6. www.vu.lt;
7. www.bodubuild.ot.lt;
8. www.lmkksc.smm.lt;
9. www.kksd.lt;
10. www.vdu.lt/sportas “Informacija apie Lietuvos sporto istorija ir pasiekimus” ;
11. ausis.gf.vu.lt;
12. www.krepsinioakademija.lt;
13. www.lass.lt;
14. www.skaters.lt “Kitoks sportas”;
15. www.sportasvisiems.lt;
16. www.aeromodeling.lt “Laisvojo skridimo, kordialinių ir radijo bangų valdomų aviamodelių varžybos”;

17. www.surask.lt/sportas “Sporto šakos. Klubai”;
18. www.lkka.lt/zurnalas/ “Ugdymas. Kūno kultūra. Sportas.”

Спорт

1. www.gazeta.ru/sport “Репортажи по всех значимых спортивных мероприятий”.
2. www.sport.com.mk;
3. www.sport-arena.com.mk;
4. www.donbass.dn.ua;
5. www.sovsport.ru “Советский спорт”, “Спортивные новости”;
6. www.rtr-sport.ru “Андрей Шевченко, жаль, что не встретимся в ЦСКА в Стамбуле”, “Ливерпуль не сможет защитить свой статус”, “Футбол”.
7. www.nevasport.ru “Все виды спорта”.

Technika ir automobilizmas

1. auto.takas.lt/straipsniai “Auto-moto”;
2. uzupis.w3.lt “Užupio patrulis”;
3. www.vkti.lt “Kelių inspekcija”;
4. www.baikeriai.lt “Lietuvos baikerių centras”;
5. www.jeep.lt “Dakaras nuo pradžių”;

Техника и автомобили

1. garant.auto.ru “Автомобили”;
2. auto.germany.ru “Виды автострахования”;
3. pravo.ruz.net “Статьи”;
4. www.dgv.pt “На дороге с большей безопасностью”;
5. www.evtushenko-spb.narod.ru “Юридический самоучитель водителя”;
6. day.udm.net “Предчувствие автогражданской войны”;
7. www.altmaster.ru “Ответственность водителя”;
8. www.autocentre.ua “Автомобили”;
9. www.truckmarket.ru “Лента новостей”

Bankininkystė

1. www.lba.lt “Apie mus”;
2. finansai.tripod.com “Centrinio banko veiklos principai ir vaidmuo ekonomikoje”, “Finansų rinkos vystymasis į bankų sistemą”, “Pinigai ir jų cirkuliacija”, “Bankų restruktūrizavimo

patirtis Vidurio Europos šalyse ir jos taikymas Lietuvoje”, “Bankinės krizės pamokos ir dabartinio bankinio sektoriaus raida Lietuvoje”;

3. www.vpk.lt “Lietuvos Respublikos vertybinių popierių komisija”;
4. www.jla.lt “Bankų konkurencija”.

Банковское дело

1. www.defakta.kz “Вы даёте задаток”;
2. www.chif.ru “Регистрация на недвижимость”;
3. nalog.akcentplus.ru “Налоговый кодекс”;
4. www.anika.ru.

Война

1. www.iraquwar.ru ;
2. docenter.krs.ru;
3. www.medialaw.ru.

Karas

1. www.kalnieciai.lt;
2. www.nato.ot.lt;
3. www.nazi.projektas.lt;
4. www.klingys.com;
5. www.xxamzius.lt;
6. www.ziemgala.lt;
7. buves.lrv.lt;
8. fc.vdu.lt.

Grožinė literatūra

1. anthologu.lms.lt.

Художественная литература

1. www.vavilon.ru;
2. www.guelman.ru .

Seimas

1. www.lrs.lt “Lietuvos respublikos seimas”;
2. www3.lrs.lt/lrs_vaikams “Seimas vaikams”;
3. www.lietuva.lt “Seimas. Konstitucija”;
4. www.on.lt “Lietuva internete”;
5. www.evaldzia.lt “Viešosios paslaugos”.

Животный мир

1. www.jelezka.ru “Последние статьи”;
2. www.minpriroda.by ;
3. www.commerce.u “Домашние животные”;
4. www.zum.lt “Племенные животные”;
5. www.animal.ru;
6. www.cat.ru;
7. www.geocities.com;
8. www.aqarium.ru.

Medicina

1. www.medicine.lt “Išsamūs resursai apie mediciną Lietuvoje”;
2. www.bpg.lt “Šeimos medicina”;
3. www.surask.lt/medicina_sveikata ;
4. medicine.dnp.lt “Liaudies medicina”;
5. www.takas.lt “Medicina”;
6. www.holistic-medicina.lt “Sielos ir proto harmonija”;
7. anonalko.tinklapis.lt “Anoniminiai alkoholikai”;
8. www.jppc.lt “Straipsniai”;
9. www.aai.lt;
10. www.psichiatrija.ot.lt;
11. www.mtp.lt;
12. www.psichoterapija.ot.lt;
13. www.unionclinic.lt;
14. ausis.gf.vu.lt;
15. www.supermama.com;
16. www.onkonet.lt;
17. www.kontracepcija.lt;
18. www.vaikas.lt.

Медицина

1. www.med2000.ru “Статьи”;
2. www.narmed.ru “Статьи”;
3. www.medlinks.ru;
4. health.mpei.ac.ru;
5. www.medland.ru;

6. www.mosmed.ru “Медицина”;
7. www.x-medicine.com/ru/ “Домашняя медицина”;
8. tibet.med.narod.ru “Тибетская медицина”;
9. agat.aorta.ru “Медицина – библиотека популярных изданий”;
10. top100.rambler.ru “Медицина”.