

Mikhailov, Mikhail 2017. «Ključevye slova»; avtorskij stil' i perevod. Scando-Slavica 63:17

<https://www.tandfonline.com/doi/full/10.1080/00806765.2017.1320109>

"Ключевые слова", авторский стиль и перевод

Abstract

Keywords, writer's style, and translation

One of the ways to detect characteristics of a writer's style is to compare frequency lists of works in question with those from a control corpus. The popular software package WordSmith Tools has a utility called Keywords, which compares frequency lists using the log-likelihood coefficient. However, WordSmith Tools can only process unlemmatized lists, and this limits the use of the program. In this paper, the keywords are searched for in lemmatized frequency lists with the help of the software package R. The focus of the research was on works by Mikhail Bulgakov and their translations into Finnish. The control data were literary works by Russian authors of the same period and their translations into Finnish. The study demonstrates that the comparison of lemmatized word lists can yield more interesting results than that of unlemmatized data. The analysis of the keywords acquired from Bulgakov's works reveals some interesting features of the writer's style, e.g. his fancy for verbs of speech. Comparing this list with the keywords of the Finnish translations of the same works shows that translators preserve only lexical features of the style. Grammatical peculiarities of the original texts reflected in frequencies of prepositions, conjunctions, particles, etc. could not be traced in translations.

Keywords: corpus-based studies, writer's style, translator's style, literary translation, quantitative analysis of style, keyword analysis, Russian-Finnish translation

1. "Ключи" к тексту?

В популярном пакете программ WordSmith Tools (далее WST)¹ имеется утилита под названием Keywords ("Ключевые слова"). Довольно многозначный термин "ключевое слово"² в данном компьютерном приложении понимается строго статистически: это текстоформа, которая встречается в исследуемых текстах чаще или реже, чем в других текстах, и эта разница является статистически значимой. Многие исследователи пользуются для исследования WST. Однако эта программа имеет ряд недостатков, о которых будет рассказано в разделе 2. Сам по себе метод технически не очень сложен и, имея некоторые навыки программирования, поиск ключевых слов можно выполнять и самостоятельно, что и будет продемонстрировано в этой статье.

Анализ ключевых слов (далее также АКС) может использоваться для решения самых разных задач. Можно выявлять особенности региональных вариантов языка, жанров, изменения в языковой норме. Например, Уилкинсон использует АКС для выявления лексических различий между британским и американским вариантами английского (Wilkinson 2014), Бель ищет с помощью АКС различия между языком юридических документов ЕС и польскими законодательными актами (Biel 2014). В работах Кемппанена (Kemppanen 2004 и 2008) и Пробирской (Probirskaja 2009) этот метод используется для поиска "идеологически нагруженных" слов. Довольно популярен АКС и для анализа языка новостей и политического дискурса (см. например, Fidler & Cvrček 2015, Johnson & Esslin 2006, Milizia 2010, Seale et al 2006). Примером использования метода для исследования лексических особенностей языка и стиля является анализ пьесы

1 Подробная информация о программе, а также ее дистрибутивы размещены на сайте <http://www.lexically.net/>. См. также Scott & Tribble 2006.

2 Назовем хотя бы "ключевые слова культуры" у Анны Вежбицкой (Anna Wierzbicka. *Understanding Cultures through Their Key Words: English, Russian, Polish, German, and Japanese.* Oxford University Press 1997).

Шекспира "Ромео и Джульетта" в книге Скотта и Триббла (Scott & Tribble 2006: 59-63). Чермакова и Фарова используют ключевые слова как критерий оценки качества переводов "Гарри Поттера" на чешский и финский языки (Cermakova & Farova 2010).

Для поиска ключевых слов нужны два текстовых массива: экспериментальный (ЭМ), то есть тексты, являющиеся объектом исследования, и контрольный (КМ), то есть тексты, представляющие некую «норму». Выбор КМ может довольно сильно повлиять на результаты исследования. Предположим, например, что требуется выявить особенности языка газеты "Аргументы и факты". Нужна ли в контрольном массиве региональная пресса, или следует ограничиться прессой федерального уровня? Должны ли в КМ кроме газет быть представлены также журналы? Являются ли новостные интернет-порталы прессой, нужно ли будет также включить в контрольный массив интернет-публикации? Фидлер и Цврчек обращают внимание на эту проблему и предлагают использовать несколько контрольных массивов, чтобы сделать результаты более объективными (Fidler & Svřek 2015: 200). Другое решение – всегда использовать в качестве контрольного массива один и тот же большой корпус литературного языка, например, в исследованиях на материале английского языка таким корпусом часто оказывается BNC, Британский национальный корпус (см., напр., Johnson & Esslin 2006).

Возможен и другой формат проведения исследования, когда объектом сравнения оказывается не экспериментальный и контрольный массивы, а два экспериментальных массива, то есть ни один из них не принимается в качестве представителя "усредненного стандарта". Примерами таких исследований может быть, например, сравнение "демократической" и "патриотической" публицистики, женской и мужской речи и т. п. Вряд ли можно утверждать, что женщина говорит

более "нормативным" языком, чем мужчина, и что язык патриотических изданий чем-то "лучше" или "хуже" языка прессы демократического лагеря.

Несмотря на то, что анализ ключевых слов, безусловно, является интересным и полезным исследовательским методом, он имеет и свои недостатки. Во-первых, он неэффективен при работе с небольшими массивами данных. Короткие частотные списки не обеспечивают достаточного объема данных для выполнения статистических тестов, и в результате программа находит лишь служебные слова, имена персонажей и топонимы. Во-вторых, для получения надежных результатов желательно, чтобы КМ был существенно больше по объему, чем ЭМ (Berber Sardinha 2000), но это далеко не всегда выполнимо. В-третьих, иногда трудно решить вопрос о составе контрольного массива, и поэтому выбор может сильно повлиять на результаты исследования.

В настоящей статье делается попытка применить анализ ключевых слов для выявления индивидуальных особенностей языка писателя. Интересным представляется также выяснить, насколько эти особенности сохраняются при переводе произведения на другой язык, и сохраняются ли они вообще. Мы продемонстрируем применение метода на примере произведений М.А. Булгакова и их переводов на финский язык.

Булгакову посвящено огромное количество литературы, однако языковыми аспектами его творчества занимаются не так активно. Отмечается диалогичность и драматургичность прозы Булгакова, обилие прямой речи Соколов (2003: 542-548), диалоги и дискуссии персонажей, чем-то даже напоминающие диалоги у античных философов (см, напр., Baker 1998). Многие исследователи находят у Булгакова реминисценции и пародирование Толстого (Vojvodić 2001), Гоголя (Ludvig 2001), а Соколов пишет о влиянии Толстого, Гоголя и других писателей

XIX века на формирование творческой манеры писателя Соколов (2003: 542-548). Читая Булгакова, нельзя не обратить внимания на его некоторую "ретроградность", симпатии к старому, которые проявляются, например, в книжности речи многих персонажей: профессора Преображенского, Воланда, Понтия Пилата. Эрик Лаурсен (Laurson 2007) пишет о смешении классической правильной речи положительных героев "Собачьего сердца" с канцелярским новоязом с просторечными элементами в речи Швондера и Шарикова. Мэй Рэйчел (Rachel 1998), сравнивая три перевода "Мастера и Маргариты" на английский язык, отмечает именно диалогичность и маркеры разговорности как главные вызовы для переводчика Булгакова на английский.

Мариэтта Чудакова (1991) в своей статье тоже упоминает о консерватизме Булгакова и его настороженном отношении к экспериментированию в литературе, что находит отражение, например, в ироничном тоне, в котором описываются футуристы и их деятельность в романе "Белая гвардия". Однако Булгаков был лично знаком с Пастернаком и в творчестве писателей прослеживаются некоторые интересные параллели. Например, романы "Белая гвардия" и "Доктор Живаго" начинаются со сцены похорон матери главного героя, встречаются похожие эпизоды, когда герой засыпает в неподходящий момент, в обоих романах постоянно идет снег.

Ни одна из найденных нами работ о языке Булгакова не оперирует количественными данными. Представляется интересным выяснить, подтвердят ли результаты нашего исследования с применением квантитативной методики результаты предшествовавших исследований и позволят ли они узнать о языке Булгакова что-либо новое.

2. Методика получения списков ключевых слов

Чтобы вычислить "ключевой характер" языковой единицы (англ. "keyness", далее: К), требуются следующие данные:

- ее частота в ЭМ,
- размер ЭМ в словоупотреблениях,
- частота единицы в КМ,
- размер КМ в словоупотреблениях.

Для вычисления значения К могут применяться разные статистические тесты, например, хи-квадрат (chi-square test, χ^2) или тест логарифмической функции правдоподобия (log-likelihood test). Тесты позволяют измерить степень наблюдаемого различия и определить, является ли оно статистически значимым. Если результаты позволяют это утверждать, единица включается в список ключевых слов.³

Таким образом, для выполнения поиска ключевых слов достаточно двух частотных списков и знания размеров текстовых массивов. Поэтому поиск может быть осуществлен и без WST. Требуемые вычисления можно, в частности, выполнить с помощью пакетов программ для статистического анализа,⁴ например R. Поиск производится следующим образом:

1. Частотные списки объединяются в сводную таблицу со столбцами "Слово", "Частота в ЭМ", "Частота в КМ".

3 Подробное описание метода см. <http://ucrel.lancs.ac.uk/llwizard.html>. Описание упомянутых статистических тестов можно также найти, например, в книге Оукса (Oakes 1998).

4 Хотя электронные таблицы Excel или Calc также располагают необходимыми возможностями, выполнение тестов с их помощью очень трудоемко.

2. В сводную таблицу добавляются столбцы с ожидаемыми значениями частот в ЭМ и КМ. Приведем формулу для расчета ожидаемой частоты слова в ЭМ:

<Формула 1>

где E_e — ожидаемая частота в ЭМ, O_e — наблюдаемая частота в ЭМ, O_c — наблюдаемая частота в КМ, N_e — размер ЭМ, N_c — размер КМ.

3. В сводную таблицу добавляется столбец с результатами статистического теста. Например, это может быть тест логарифмической функции правдоподобия (log-likelihood), который рассчитывается по формуле:⁵

<Формула 2>

Полученный коэффициент всегда больше нуля, но для того, чтобы различать случаи, когда слово имеет более высокую относительную частоту в ЭМ и когда оно частотнее в КМ, во втором случае коэффициенту присваивается знак "минус".

4. Уничтожаются строки с низкими значениями K и с частотами слов ниже заданной. Оставшиеся строки упорядочиваются по убыванию K .

В приложении 2 к настоящей статье приведены команды R, с помощью которых можно получить списки ключевых слов.

Следует подчеркнуть, что полученный список является лишь списком слов, которые встречаются в исследуемых текстах чаще или реже "нормы". Причины отклонений могут быть различными, например, тематика текстов, жанр, региональные особенности, идиостиль автора. Поэтому релевантность или нерелевантность тех или иных элементов списка зависит от задач исследования.

5 В литературе встречаются разные варианты формулы. Мы использовали метод расчета, описанный Полом Рэйсоном на сайте Ланкастерского университета (Великобритания) (<http://ucrel.lancs.ac.uk/llwizard.html>).

Поиск ключевых слов можно выполнять как на нелемматизированных, так и на лемматизированных частотных списках. Результаты могут довольно существенно отличаться, особенно в языках с богатой морфологией.

У WST очень ограниченные возможности работы с аннотированными текстами, а предлагаемая лемматизация крайне примитивна (предлагается задавать списки форм для каждой леммы, что проблематично даже для английского языка из-за морфологической омонимии), поэтому с помощью этого пакета можно получать списки ключевых слов лишь из нелемматизированных частотных списков. Однако, пользуясь методикой, описанной выше, можно самостоятельно получить списки ключевых слов из лемматизированных словарей, пользуясь пакетом программ R. В данной статье мы сравним списки, полученные из нелемматизированных и лемматизированных словарей, и выясним, улучшает ли лемматизация результаты поиска.

2. Экспериментальный и контрольный массивы. Получение списков ключевых слов

Для анализа были выбраны три произведения Михаила Булгакова, написанные в разные периоды творчества писателя: "Собачье сердце" (1925), "Мастер и Маргарита" (1929-1940) и "Театральный роман" (1936-1937). Произведения переведены на финский язык разными переводчиками: "Театральный роман" и "Собачье сердце" перевел Эса Адриан (1971 и 1975), "Мастера и Маргариту" – Улла-Лииса Хейно (1969) (список текстов экспериментального массива см. также в приложении 1). Таким образом, в финском языковом материале представлены два переводчика, и это позволит судить именно об отражении стиля писателя в переводе, а не о "почерке" переводчика.

Для контрольного массива были выбраны произведения русских писателей разных литературных направлений, относившихся к тому же поколению, что Булгаков, и опубликованные в тот же временной период (1920-40-е гг.). Это произведения И. Бабеля, А. Фадеева, И. Ильфа и Е. Петрова, Ю. Олеши, Б. Пастернака, М. Шолохова и М. Зощенко и их переводы на финский язык (полный список тестов контрольного массива см. в приложении 1). Полученные контрольные массивы оказались примерно в два раза больше по количеству словоупотреблений (с/у), чем экспериментальные (см. табл. 1), что, как уже отмечалось выше, повышает надежность результатов.⁶ Тексты, включенные в контрольный массив, являются достаточно разнообразными по тематике и стилю, и переведены на финский язык разными переводчиками⁷, в их числе – Адриан и Хейно, переводы которых вошли в ЭМ.

Лемматизация русских текстов была выполнена с помощью программы R Morph (Gelbukh & Sidorov 2003), часть грамматической омонимии была снята с помощью программ, написанных автором настоящей статьи, в некоторых текстах омонимия была снята вручную. Финские тексты были лемматизированы с помощью программы Connexor Machine Syntax⁷, часть повторяющихся ошибок была снята списком.

Таблица 1. Объемы экспериментальных и контрольных текстовых массивов

<Таблица 1>

6 Различия в размерах русских и финских субкорпусов связаны с межъязыковыми грамматическими различиями (например, в финском языке очень мало предлогов и довольно много сложных слов). Обычно тексты на русском языке «длиннее» финских. В данном исследовании финский контрольный массив оказался больше по объему лишь из-за того, что в корпус было включено несколько разных переводов некоторых произведений.

7 <http://www.ling.helsinki.fi/atk/sovellusohj/parsers/fi-fdg.shtml>

Списки ключевых слов были получены двумя способами:

1. С помощью WordSmith Tools. Полученные ключевые слова — текстоформы.
2. Путем обработки лемматизированных словников с помощью программы R. Полученные ключевые слова — леммы.

Сравним результаты поиска по нелемматизированным частотным спискам с результатами обработки лемматизированных списков. В таблице 2 приведены первые 20 слов списка, полученного с помощью WST, в таблице 3 — первые 20 слов из лемматизированного списка.

Таблица 2. Первые 20 кандидатов в ключевые слова, полученных с помощью WordSmith Tools

<Таблица 2>

Таблица 3. Первые 20 кандидатов в ключевые слова, полученных путем обработки лемматизированных списков

<Таблица 3>

Формула 2 действует только для случаев, когда частота слова в обоих массивах больше нуля, поскольку $\ln(0) = -\infty$. В WST, по-видимому, используется специальный вариант формулы для случаев, когда слово не встретилось в одном из массивов. Наши расчеты производились строго по формуле 2, возвращающей

для таких случаев ошибку, и поэтому в табл. 3 нет слов с нулевой частотой в каком-либо из массивов, тогда как в табл. 2 таковых одиннадцать (то есть более половины!). Отметим, что во многих случаях элементы, не встретившиеся в одном из списков, интереса не представляют, и их исключение только улучшает качество результатов.

Сравним списки, приведенные в табл. 2 и 3. В табл. 2 более половины списка составляют имена собственные (12), 16 являются существительными. В списке всего один глагол (*ответил*) и одно наречие (*совершенно*). В табл. 3 всего два имени собственных: *Иван* и *Римский* (фамилия персонажа "Мастера и Маргариты"), в списке 11 существительных (включая фамилию *Римский*, которую лемматизатор ошибочно отнес к прилагательным), три глагола (*ответить*, *отозваться* и *заговорить*), и есть одно прилагательное *совершенный* (в которое лемматизатор, возможно, включил, кроме форм прилагательного *совершенный*, также и наречие *совершенно*).

Анализ полных списков показывает, что в списке ключевых слов, полученном из нелемматизированного списка, почти 70% составляют имена существительные и имена собственные, глаголов чуть больше 9%, наречий — 6,3%, а прилагательных — 1,66%. Ключевые слова из лемматизированного списка намного разнообразнее: существительных около 40% (при этом имен собственных — меньше трех процентов⁸), глаголов — более 20%, прилагательных — 14%, а наречий — 8%. Несмотря на неизбежность ошибок при лемматизации, есть все основания считать, что поиск ключевых слов по лемматизированным спискам дает намного более интересные результаты для языков с богатой морфологией. Причиной этого является то, что в нелемматизированном списке частотность слов "распыляется"

8 Выше было уже отмечено, что имена, не встретившиеся в одном из текстов, не включаются в список. Русский лемматизатор не определяет лемму для слов, отсутствующих в словаре системы, поэтому редкие имена и фамилии отсутствуют в списках изначально.

по разным грамматическим формам (например: *играть* → *играю, играешь, играет, ... играл, играла, играло, играли, ... играя, ...* и т. д.), и в итоге заметными остаются только неизменяемые слова или слова, часто употребляющиеся в какой-либо одной форме.

Таким образом, представляется целесообразным выполнять дальнейшую работу на лемматизированных русских и финских частотных списках. Ранее автором статьи уже выполнялся анализ данного материала, но с помощью WordSmith Tools и на нелемматизированных текстах (см. Mikhailov, Cooper 2016: 133-144). Хотя некоторые из явлений, о которых будет идти речь в настоящей статье, удалось выявить и с помощью нелемматизированных списков, привлечение лемматизации позволило сделать целый ряд новых наблюдений, а также сделать полученные в ходе предыдущего исследования результаты более убедительными.

При составлении списка ключевых слов исследователь задает нижний предел значения коэффициента K . Кроме того, стоит учитывать и абсолютное значение частоты слова, ведь низкочастотные слова нельзя назвать ключевыми по определению. В данном исследовании в список ключевых слов включались слова со значением K более 10 или менее -10 ("антиключевые" слова) с абсолютной частотой более 5 хотя бы в одном из текстовых массивов. В результате были получены русский список из 532 слов и финский список из 890 слов. Интересно, что русские и финские лемматизированные частотные списки, по которым выполнялся анализ, были ближе по размерам: русские — 15 800 (ЭМ) и 31 000 (КМ), финские — 16 300 и 30 500. Это может быть признаком того, что между переводами больше лексических отличий, чем между оригиналами, с которых делался перевод. Но это противоречит существующим представлениям о стандартизации языка при переводе (см., например, Olohan 2004: 91).

3. Ключевые слова Булгакова

Итак, полученные нами списки состоят из слов ЭМ, частоты которых сильно отличаются от частот тех же слов в КМ. Далеко не все они связаны с идиостилем автора, значительная часть найденных слов определяется тематикой текста.

При изучении списков в первую очередь обращалось внимание на слова с более высокой частотой и на слова, образующие группы. Просмотр списков производился по частям речи.

Совершенно очевидно, что **имена собственные** почти гарантированно оказываются в списках ключевых слов и при этом мало связаны с индивидуальным стилем автора. Среди **нарицательных существительных** полученного нами списка также не удалось найти слова, которые можно было бы назвать маркерами индивидуального стиля. Список существительных позволяет узнать, о чем пишет автор, где происходит действие, и как зовут персонажей. В полученном нами списке ключевых слов встретились следующие существительные: *роман, актер, квартира, профессор, поэт, дворец, голова, трамвай, кот*. Все они связаны с содержанием текстов, включенных в наш ЭМ. Нам не удалось выявить увлечения автора существительными какого-либо определенного грамматического или семантического класса. Это, впрочем не означает, что на существительные не следует обращать внимания, ведь некоторые авторы могут активно употреблять окрашенную лексику, архаизмы, диалектизмы, просторечие, – и эти особенности могут найти отражение и в употреблении имен существительных.

Прилагательные как слова, с помощью которых автор дает более подробную информацию о предметах, являются важным элементом описаний. В нашем списке ключевых слов только два цветообозначения — *черный* и *рыжий*, три

прилагательных, обозначающих размеры — *громадный, большой и огромный* (два последних являются антиключевыми), и полностью отсутствуют прилагательные запаха и вкуса. Это означает, что частотность последних у Булгакова в целом не отличается от частотности их у других авторов его времени. Интересным представляется то, что в списке много прилагательных со значением высокой степени признака: *совершенный, громадный, чрезвычайный, необыкновенный* и др., всего 13. Среди слов с отрицательным значением *K* много высокочастотных прилагательных: *новый, старый, большой, остальной* и др. Разница в частотах между ЭМ и КМ оказывается довольно заметной, например, слово *новый* встречается у Булгакова 70 раз (0,39/1000), а в КМ — 371 раз (1,09/1000).⁹ Вероятно, Булгаков, избегая банальности, старался больше употреблять эмоционально яркие прилагательные — *громадный, беззвучный, отчетливый*, — или не употреблять никаких.

В списке ключевых слов около 20% **глаголов**, из которых значительная часть является глаголами речи, причем именно у глаголов речи наиболее высокое значение *K*. Первые шесть глаголов списка: *ответить, отозваться, заговорить, вскричать, воскликнуть*. Обилие глаголов речи ощущается даже в списке ключевых слов, полученном из нелемматизированного словника, но в лемматизированном списке эта тенденция становится еще более заметна. Это указывает как на то, что в текстах Булгакова много прямой речи, так и на то, что писатель придает большое внимание описанию обстоятельств речевого общения. Кроме нейтральных общеупотребительных глаголов *спросить, сказать, ответить* (которые, тем не менее, все попали в наш список) Булгаков использует окрашенные слова, такие как *шепнуть, бормотать, умолять* и т. п. В списке не

9 Напомним, что КМ в два раза больше, чем ЭМ, поэтому относительные частоты отличаются не так сильно.

встретилось глаголов речи с отрицательным значением *K*, то есть в этой группе нет слов, которых Булгаков сознательно избегал бы.

Другая выявленная нами особенность употребления глаголов в произведениях Булгакова — доминирование глаголов совершенного вида. Среди слов с положительным значением *K* 78 глаголов совершенного вида и лишь 18 — несовершенного. При этом среди антиключевых слов наблюдается противоположная тенденция — 16 глаголов несовершенного и только 2 совершенного вида. Это может говорить о том, что в текстах Булгакова описания событий преобладают над описаниями состояний и процессов. Для проверки этого утверждения требуется исследование макроструктур текстов корпуса.

Среди **наречий**, попавших в список ключевых слов, вновь отметим слова со значением высокой степени признака: *совершенно, чрезвычайно, очевидно, очень*. Слово *совершенно* является «любимчиком» Булгакова: оно является частотным во всех трех текстах («Мастер и Маргарита»: 141 (1,25 на 1000 с/у), «Собачье сердце»: 41 (1,67/1000), «Театральный роман»: 26 (0,63/1000)). Правда, в «Театральном романе» его частота заметно ниже, чем в первых двух. Возможно, писатель заметил, что употребляет это слово слишком часто, и начал его избегать. Для сравнения, в романе А. Фадеева «Разгром» слово *совершенно* употреблено 6 раз, относительная частота 0,14/1000. Наречие *опять* попало в список ключевых слов, в то время как его синоним *снова* является антиключевым, к этой паре мы вернемся в разделе 4 настоящей статьи.

Нередко при выполнении анализа по ключевым словам из рассмотрения исключаются **служебные слова**, поскольку они мало связаны с тематикой текстов. В программе WST даже есть возможность задать «стоп-словарь», т. е. список слов, которые будут игнорироваться при анализе. В нашем исследовании все

происходит наоборот: игнорируются слова, связанные с тематикой, а слова, по которым можно судить о лексических и грамматических предпочтениях автора — в том числе предлоги, союзы, частицы, — становятся предметом рассмотрения.

Среди служебных слов более высокая частота по сравнению с контрольным массивом у слов¹⁰ *что, и, а*, и в несколько меньшей степени — *но*. Для предлога *в* значение *K* довольно высокое: 82,89, при этом два других предлога — *но* и *для* — оказались среди «антиключевых» слов с *K* равным -64,52 и -61,02 соответственно. Отклонения в частотности употребления предлогов могут быть связаны с тем, что автор любит или не любит употреблять некоторые стандартные обороты.

Более высокая относительная частота предлога *в*, по всей вероятности, связана с более высокой частотой локативных конструкций (*в доме, в дом*), поскольку каких-либо устойчивых лексических коллокаций с этим предлогом обнаружить не удалось. Что касается предлога *но*, то в анализированных текстах Булгакова очень редко по сравнению с КМ употребляются обороты *по случаю, по сторонам, по пути*. Еще интереснее оказалась ситуация с употреблением предлога *для*: в анализированных текстах Булгакова, как оказалось, почти не употребляются конструкции с местоимениями, такие как *для меня, для вас, для этого* и т.п. Эти особенности были выявлены с помощью построения списков коллокаций.

В список попали и многие частицы: *же, уж, да, лишь, ни, ну*. Все эти слова намного частотнее у Булгакова, например, относительная частота употребления частицы *же* 5,16 на 1000 с/у, а в КМ — лишь 2,91. Все эти слова являются маркерами живой разговорной речи. Их высокая частотность является, таким образом, признаком диалогичности и сценичности булгаковской прозы, о которой говорилось в начале данной статьи.

10 В настоящей версии аннотации корпуса пока не снята омонимия союзов и частиц *и* и *а*, а также союза *что* и вопросительного местоимения *что*.

4. Ключевые слова в финских переводах

Напомним, что финский список ключевых слов был получен путем сравнения лемматизированных частотных списков переводов произведений Булгакова и переводов произведений, вошедших в русский контрольный массив. Как известно, финский язык относится к другой языковой семье, и его грамматика сильно отличается от русской: полностью отсутствует категория рода, сильно отличается система глагольных времен, нет категории глагольного вида. Посмотрим, удалось ли переводчикам сохранить особенности булгаковского стиля и насколько особенности лексики и грамматики повлияли на полученные нами списки ключевых слов.

При визуальном изучении существительных финского списка ключевых слов возникает впечатление, будто русский список "перевели" на финский: в нем очень много словарных соответствий слов русского списка. В таблице 4 сведены первые двадцать существительных из русского и финского списков. Легко убедиться, что в этих списках много общего.

Таблица 4. Первые 20 существительных из русского и финского списков ключевых слов

<Таблица 4>

Лишь для шести слов из русского списка нет соответствий в финском списке и только четыре слова из финского списка не соответствуют какому-либо слову из русского списка (слова, не имеющие соответствий, выделены жирным шрифтом).

При этом, у русских слов *гость*, *луна*, *голос* и *дверь* и у финских слов *puhelin*

'телефон' и *miliisi* 'милиция' есть в противоположных списках соответствия, не попавшие в первую двадцатку.

Любовь Булгакова к глаголам речи, которую мы заметили, изучая списки ключевых слов, полученных из русских текстов, видна и в текстах финских переводов. Один финский глагол, в силу вышеупомянутых различий в грамматической системе, может соответствовать нескольким русским глаголам. Тем не менее, в русском списке 21 глагол речи, а в финском их лишь немногим меньше: 18. В таблице 5 сопоставлены глаголы речи из русского и финского списков ключевых слов. Видовые пары и синонимы были объединены в группы. Оказалось, что значительная часть русских глаголов списка имеет финские соответствия, у некоторых групп даже суммы частот оказались довольно близкими, например, *ответить/отозваться* (404 + 103 = 503) — *vastata* (541), *шептать/шепнуть* (47 + 49 = 106) — *kuiskata* (104). Доминирование глаголов совершенного вида, замеченное в русском списке ключевых слов, не находит отражения в финском списке, поскольку разным глаголам русской видовой пары обычно соответствует один финский глагол, например, *читать / прочитать* — *lukea*, *писать / написать* — *kirjoittaa*. Интересно, что в финском списке ключевых слов существенно больше глаголов, чем в русском (231 против 119), несмотря на то, что в финском языке нет категории вида и богатой глагольной префиксации. Причины этого явления требуют дальнейшего изучения.

Таблица 5. Глаголы речи из русского и финского списков ключевых слов

<Таблица 5>

В финском списке, как и в русском, много прилагательных и наречий со значением высокой степени признака. Так же, как и в русском списке, нейтральные высокочастотные прилагательные являются "антиключевыми" словами, даже несмотря на то, что многие из этих слов сильно отличаются по семантике от своих русских соответствий, например, прилагательное *nuori* означает 'новый' и 'молодой', *pitkä* — 'длинный', 'долгий', и 'высокий'.

Соответствия многих наречий, оказавшихся ключевыми словами Булгакова, также удается найти в финском списке ключевых слов.

Таблица 6 Наречия в русском и финском списках ключевых слов

<Таблица 6>

Как видно из таблицы 6, у некоторых пар частота оказывается очень близкой, например, *опять* / *taas*, *часто* / *usein*. Но гораздо чаще оказывается, что частоты русских слов и их финских соответствий различаются очень сильно. Например, слово *затем* встретилось в исследовательском массиве 129 раз, а его финский аналог *sitten* употреблен в финских переводах тех же текстов 748 раз. На самом деле финские переводчики последовательно употребляли *sitten* в качестве соответствия для русского *затем*, но это же слово оказывается соответствием для многих других русских устойчивых сочетаний: *потом*, *(тому) назад*, *после*, *в таком случае*.

Другая пара — *тотчас* / *heti* — демонстрирует ту же тенденцию: частота русского ключевого слова 89, в то время как финское соответствие встретилось в переводах

229 раз. Причины те же: *heti* является также соответствием для слов *сразу*, *моментально*, *немедленно*, *сейчас*, *тут же*.

Выше уже упоминались наречия *опять* и *снова*. Эти два синонима довольно близки по частоте в русском контрольном массиве (150 и 216)¹¹. В языке Булгакова слово *опять* является ключевым: оно встретилось 163 раза и относительная частота почти в два раза выше, чем в контрольном массиве. Слово *снова*, напротив, является "нелюбимым" и встретилось всего 5 раз, в результате чего оказалось в списке антиключевых слов. Следует отметить, что в финском языке довольно много слов и выражений со значением 'опять': *taas*, *jälleen*, *uudelleen*, *uudestaan*, *vielä*. Поэтому высокая частота слова *taas* в финских переводах может показаться странной. Это объясняется тем, что слово *taas*, так же, как и упоминавшиеся выше *sitten* и *heti*, употребляются во многих других функциях. Действительно, анализ конкорданса показывает, что *taas* может быть эквивалентом для русских союзов *а*, *и*, *но* и частицы *же*. Все эти слова, как уже говорилось выше, являются ключевыми словами Булгакова. Вероятно, именно этим и объясняется высокое значение К для слова *taas*. А близость частот слов *опять* в оригинальных текстах (163 и 150) и *taas* в переводах (140 и 142) скорее всего является совпадением.

Просмотр конкорданса показывает, что в качестве эквивалента для русского *опять* далеко не всегда используется *taas*, слово *jälleen* оказывается по крайней мере не менее частотным:

— Молчать! — вскричал Пилат и бешеным взором проводил ласточку, опять впорхнувшую на балкон.

— Vaiti! Pilatus huusi ja saattoi raivokkain katsein pääskyä, joka jälleen pyrähti parvekkeelle. ("Мастер и Маргарита", пер. У.-Л. Хейно)

¹¹ Отметим, что в наших материалах наречие *снова* более частотно у авторов первой половины 20 века, но, к концу 20 века более частотным становится наречие *опять*.

Продемонстрированные выше данные показывают, что нельзя делать заключения, опираясь исключительно на статистику. Нужно обязательно хотя бы выборочно проверять и контексты. Особенно трудно судить об употреблении служебных слов: оно тесно связано со структурой текста, с наличием в тексте сложных конструкций, оборотов, перечислений и т. д. Причем, если при переводе с одного языка на другой слова знаменательных частей речи находят в переводе непосредственное отражение (то есть в обычных случаях можно найти соответствие для данного слова в переводе данного сегмента), то употребление служебных слов отражается в переводе лишь опосредованно.

Таблица 7. Служебные слова в русском и финском списках слов-кандидатов

<Таблица 7>

Союз *и* и его финское соответствие *ja* оказались единственными словами с близкой частотой. Действительно, финский союз во многих случаях оказывается соответствием русского союза, например:

Абрикосовая дала обильную желтую пену, и в воздухе запахло парикмахерской. / Aprikoosimehu kuohahti runsaana keltaisena vaahtona, ja ilmaan levisi parturin lemu.

(Мастер и Маргарита, пер. У.-Л. Хейно)

И тут надо мною склонилось широкое лицо с круглейшими очками. / Ja silloin ylitseni kumartuivat leveät kasvot ja pyöreät silmälasit.

(Театральный роман, пер. Э. Адриан)

Но соответствие *и* → *ja* является распространенным только тогда, когда *и* функционирует в качестве союза.

могут появляться вместо союза *ja* различные клитические частицы *-kin/-kaan/-kään/-han/-hän/-pa/-pä*:

Эту поэму Иван Николаевич сочинил, и в очень короткий срок...

Tämän runoelman Ivan Nikolajevitš olikin sepittänyt, vielärä hyvin lyhyessä ajassa...

(Мастер и Маргарита, пер.У.-Л.Хейно)

Что касается слов *что* и *että*, то частота первого почти в два раза превышает частоту последнего. Это в первую очередь связано с тем, что кроме союза *что* есть и местоимение *что*, которому соответствуют финское местоимение *mikä*.

Различие в частотах слов *впрочем* и *sentään* (81 и 42) так же, как и в предыдущих случаях, связано с тем, что у слова *впрочем*, кроме *sentään*, есть еще несколько соответствий, например *kuitenkin, muuten, toisaalta*.

Различия между частотами употребления предлога *для* и послелога *vuoksi* еще более заметны (126 и 39). Дело в том, что предлоги намного более употребительны в русском языке, чем в финском, в котором грамматические значения намного чаще выражаются с помощью падежных форм без предлога. Например, в конструкции *для + Р. п.* где объект является лицом, употребляется форма аллатива, например, *подарок для Анны / lahja Annalle*. Кроме того, послелогу *vuoksi* синонимичны послелоги *varten* и *hyväksi*. Это также приводит к снижению частотности финского соответствия.

Разница в частотах между частицами *кроме* и *paitsi* менее значительна (110 и 87), при этом оказывается, что *paitsi* употребляется в качестве соответствия *кроме* только для контекстов *кроме того — sitä paitsi*. Другое распространенное соответствие *кроме, lukuun ottamatta*, не содержит слова *paitsi*. Вдобавок, в текстах встречается множество соответствий, не указанных в словарях, например, *kuin, vielä, ohella, muu*. В свою очередь, финское слово *paitsi* употребляется в качестве

соответствия для таких русских выражений, как *не только ... но и, к тому же, за исключением* и др. Выражение *кроме того* встретилось в текстах Булгакова 63 раза, что и обусловило близость частот слов *кроме* и *paitsi*.

5. Выводы

Наше исследование демонстрирует, что сравнение частотных списков исследовательского и контрольного текстовых массивов позволяет обнаруживать некоторые особенности авторского стиля. Лексические предпочтения находят прямое отражение, о грамматических особенностях можно судить по употреблению служебных слов. Метод ключевых слов позволяет ускорить эту работу и проверить статистическую значимость обнаруженных различий. Тем не менее, полученный список является лишь «информацией к размышлению», а не конечным продуктом.

Сравнение частотных списков, полученных из оригинальных текстов и из переводов этих текстов, показало, что лексические предпочтения оригинального текста находят отражение в переводе. Передача происходит в несколько ослабленном виде в связи с различиями в лексиконах исходного и переводящего языков.

Наблюдения над частотами служебных слов показывают, что грамматические особенности исходного текста передаются намного слабее. Это подтверждает тезис о том, что переводчик «переводит» слова текста, а грамматику текста он может только имитировать. Данная имитация может заключаться в стремлении писать такими же длинными (или короткими) предложениями, употреблять безличные или пассивные конструкции, копировать пунктуацию оригинала.

Микросинтаксис же меняется радикально, и это видно в статистике употребления служебных слов в оригиналах и переводах.

Результаты настоящего исследования еще раз подтверждают представления об уникальности авторского стиля, а также то, что передача стилистических особенностей художественного произведения при переводе на другой язык возможна лишь отчасти. О достоинствах языка оригинала некорректно судить по переводу, хотя это и практикуется.

При выявлении языковых особенностей текста через списки ключевых слов следует обращать внимание в первую очередь на слова, не связанные непосредственно с тематикой текста. Факты, полученные таким образом, требуют подтверждения и уточнения с помощью других инструментов. Например, списки коллокаций могут помочь исследовать сочетаемость кандидатов в ключевые слова, а конкордансы позволят проверить примеры употребления этих слов.

Литература

- Соколов Б. В. 2003. *Булгаков. Энциклопедия*. Москва: Алгоритм.
- Чудакова М. О. 1991. Пастернак и Булгаков: рубеж двух литературных циклов. *Литературное обозрение*. № 4, 11-17.
- Baker H. D. 1998. Socratic, Hermetic, and Internally Convincing Dialogue: Types of Interlocution in Bulgakov's the Master and Margarita. *The Russian Review*, Vol. 57, # 1, pp. 53-71.
- Berber Sardinha T. 2000. "Comparing corpora with WordSmith Tools: How large must the reference corpus be?" In *The Workshop on Comparing Corpora*. Hong Kong, Association for Computational Linguistics, pp. 7-13. <http://www.aclweb.org/anthology/W00-0902>.
- Biel L. 2014. *Lost in the Eurofog: The Textual Fit of Translated Law*. Peter Lang.

- Cermakova A. & Farova L. 2010. Keywords in Harry Potter and their Czech and Finnish translation equivalents. In Cermak F. et al (eds.). *InterCorp: Exploring a Multilingual Corpus*. Prague: Nakladatelství Lidové Noviny/Czech National Corpus, pp 177-188.
- Fidler M. and Cvrček V. 2015. A Data-Driven Analysis of Reader Viewpoints: Reconstructing the Historical Reader Using Keyword Analysis. *Journal of Slavic Linguistics*. 23(2), pp. 197–239.
- Gelbukh, A. and G. Sidorov. 2003. "Approach to construction of automatic morphological analysis systems for inflective languages with little effort." In *Computational Linguistics and Intelligent Text Processing (CICLing-2003)*, Lecture Notes in Computer Science, # 2588, Springer-Verlag, pp. 215–220. <www.cic.ipn.mx/~sidorov/GelbukhSidorovMorphCICLING2003.ps>.
- Gries S. 2009. *Quantitative corpus linguistics with R. A practical introduction*. Routledge,
- Johnson, S. & Esslin A. 2006. Language in the news: Some reflections on keyword analysis using Wordsmith Tools and the BNC. *Leeds Working Papers in Linguistics*, # 11. <http://www.leeds.ac.uk/arts/info/125154/leeds_working_papers_in_linguistics_and_phonetics/1949/volume_11_2006>
- Kemppanen H. 2004. Keywords and Ideology in Translated History Texts: A Corpus-based Analysis. *Across Languages and Cultures* 5 (1), 89-106.
- Milizia D., 2010. "Keywords and phrases in political speeches." In M. Bondi & M. Scott (eds.) *Keyness in Text*. Amsterdam & Philadelphia: John Benjamins, pp. 127-145.
- Kemppanen H. 2008. *Avainsanoja ja ideologiaa: käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi*. University of Joensuu, Joensuu.
- Ludvig S. 2001. Пародия и сатира в рассказе Булгакова 'Похождения Чичикова'. *Russian Literature*. Vol. 49, Issue 2, 15, pp. 167-182.
- Laursen E. 2007. Bad Words Are Not Allowed!: Language and Transformation in Mikhail Bulgakov's "Heart of a Dog" *The Slavic and East European Journal*. Vol. 51, # 3, pp. 491-513.

- May R. 1998. Three Translations of the Master and Margarita. *Translation Review*, 55:1, 29-33, DOI: <10.1080/07374836.1998.10523720>.
- Mikhailov M. and Cooper R. 2016. *Corpus Linguistics for Translation and Literary Studies. A guide for research*. Routledge: London and New York.
- Oakes M. 1998. *Statistics in Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Olohan M. 2004. *Introducing Corpora in Translation Studies*. Routledge: London and New York.
- Probirskaja S. 2009. *Rajankäyntiä: Suomen ja Venäjän kahdenväliset valtiosopimukset käännöstieteellisen avainsana-analyysin valossa*. Tampere University Press: Tampere.
- Scott M. & Tribble C. 2006. *Textual Patterns : Key Words and Corpus Analysis in Language Education*. Amsterdam : John Benjamins.
- Seale C, Charteris-Black J, Ziebland S. 2006. Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. *Social Science and Medicine*, 62, 10: 2577-2590.
- Vojvodić J. 2001. Один аспект пародии на Льва Толстого в ‘Записках юного врача’ М. Булгакова. *Russian Literature*, Vol. 49, Issue 2, pp. 223-233.
- Wilkinson M. 2014. Using the Keyword Tool to Explore Lexical Differences between British and American English in Specialised Corpora. *CALL-EJ* 15(1), 21-38. <http://callej.org/journal/15-1/Wilkinson_2014.pdf>

Приложение 1. Исследовательский материал

Экспериментальный массив

- Булгаков М.А., Мастер и Маргарита, 1940. Финский перевод: Saatana saaruu Moskovaan, Heino U.-L., 1969.
- Булгаков М.А., Собачье сердце, 1925. Финский перевод: Koiran sydän, Adrian E., 1975.

Булгаков М.А., Театральный роман, 1937. Финский перевод: Teatteriromaani, Adrian E., 1971.

Контрольный массив

Бабель И., Иисусов грех, 1924. Финские переводы: Jeesuksen synti, Adrian E., 1969; Jeesuksen synti, Heino U.-L., 1970.

Бабель И., Король, 1920. Финские переводы: Kuningas, Adrian E., 1950; Kuningas, Heino U.-L., 1970.

Зощенко М., Аристократка, 1923. Финский перевод: Aristokraatti, Losowitch K., 1990.

Зощенко М., Нервные люди, 1928. Финский перевод: Kireähermoista väkeä, Losowitch K., 1990.

Зощенко М., Ночное происшествие, 1940. Финский перевод: Tapaus yöllä, Losowitch K., 1990.

Ильф И., Петров Е., Золотой теленок, 1931. Финский перевод: Kultainen vasikka, Losowitch K., 1957.

Олеша Ю., Зависть, 1927. Финский перевод: Kateus, Adrian E., 1967.

Пастернак Б.Л., Доктор Живаго, 1955. Финский перевод: Tohtori Živago, Konkka J., 1958.

Фадеев А., Разгром, 1927. Финский перевод: Tuho, Heino U.-L., 1977.

Шолохов М.А., Тихий Дон, т.1, ч. 1, 1932. Финские переводы: Hiljainen Don, Siiprainen E., 1945; Hiljaa virtaa Don, Konkka J., 1956.

Приложение 2. Получение списка ключевых слов из двух

лемматизированных списков с помощью пакета программ R

R является некоммерческим приложением, работающим в различных операционных системах. Дистрибутивы доступны на сайте <https://www.r-project.org>, там же можно найти руководства и ссылки на литературу. В настоящей статье команды R приводятся с минимальными комментариями.

1. Импортировать в R в виде фреймов (*a* и *b*) две таблицы с лемматизированными списками. В каждом фрейме должны быть следующие столбцы: слово, абсолютная частота, относительная частота.

2. Объединить фреймы в фрейм *results* по полю "слово" (в нашем примере Word)

```
results <- merge(a, b, by="Word", all=TRUE)
```

3. Переименовать имена столбцов в фрейме

```
colnames(results) <- c("Word", "Exprt", "ExperRel", "Control", "ControlRel")
```

4. В полученной таблице заменить значение NA ("не задано") на 0

```
results[is.na(results)] <- 0
```

5. Добавить столбец Sign со знаком "+" (слово более частотно в ЭМ) или "-" (слово более частотно в КМ)

```
results <- transform(results, Sign = ifelse((results$ExperRel - results$ControlRel) < 0, '-', '+'))
```

6. Добавить столбцы с ожидаемым значением частоты в ЭМ (ExperE) и КМ (ControlE). Результат округляется до сотых долей. Ниже приведена команда для расчета ExperE, в формулу следует подставить значения размеров ЭМ (*size_exper*) и КМ (*size_control*).

```
result$ExperE = round((result$Exper + result$Control) * size_exper / (size_exper + size_control), 2)
```

7. Добавить столбец со значением функции log-likelihood (LL)

```
result$LL <- round((2*((result$Exper * log (result$Exper / result$ExperE)) +
(result$Control * log (result$Control / result$ControlE))))),2)
```

"Ключевые слова", авторский стиль и перевод

Таблицы и формулы

Формула 1

$$E_e = \frac{(O_e + O_c) * N_e}{N_e + N_c}$$

Формула 2

$$\frac{O_e}{E_e} = c \left(\frac{O_c}{E_c} \right)$$

Таблица 1. Объемы экспериментальных и контрольных текстовых массивов

	Русские оригиналы (с/у)	Финские переводы (с/у)
Экспериментальный массив: Булгаков	178 522	165 515
Контрольный массив: русская литература 1920-1950 гг.	340 484	348 800

Таблица 2. Первые 20 кандидатов в ключевые слова, полученных с помощью WordSmith Tools

Слово-кандидат	Частота в ЭМ	Частота в КМ	К
Магарита	422	0	895,69
Филипп	298	0	632,37
Филиппович	295	0	626
Коровьев	189	0	400,99
Воланд	186	0	394,62
прокуратор	179	0	379,77
Иван	382	118	364,47
кот	180	2	361,57
Азazelло	152	0	322,47
Пилат	137	0	290,64
Борменталь	121	0	256,69
Васильевич	124	1	252,26
ответил	331	140	248,09
что	2921	3800	227,04
Шариков	107	0	226,98
то	1565	1828	191,44
профессор	129	15	190,18
Никанор	88	0	186,67
совершенно	208	73	181,39
пес	96	3	179,31

Таблица 3. Первые 20 кандидатов в ключевые слова, полученных при обработке лемматизированных списков

Слово-кандидат	Частота в ЭМ	Частота в КМ	К
иван:N	563	153	587,55
кот:N	266	8	502,12
что:Pron	2859	3427	328,33
профессор:N	193	21	292,22
ответить	404	187	282,13
мастер:N	167	16	261,33
тот:Pron	1079	1046	239,36
театр:N	164	25	223,39
глаз:N	580	431	221,71
римский:Adj	125	9	208,35
квартира:N	225	84	189,4
я:Pron	3694	5308	170,85
отозваться: V	103	9	164,76
совершенный:Adj	215	90	164,66
же:Particle	922	990	154,26
сцена:N	127	28	148,22
роман:N	105	19	133,89
этот:Pron	2379	3322	131,72
поэт:N	79	6	130,29
заговорить: V	137	48	121,01

Таблица 4. Первые 20 существительных из русского и финского списков
КЛЮЧЕВЫХ СЛОВ

Русский	К	Финский	К
иван	587,55	ivan 'Иван'	552,01
кот	502,12	kissa 'кот'	475,38
профессор	292,22	professori 'профессор'	330,21
мастер	261,33	mestari 'мастер'	329,84
театр	223,39	näytelmä 'пьеса'	257,47
глаз	221,71	teatteri 'театр'	245,7
квартира	189,4	koira 'собака'	226,16
сцена	148,22	romaanі 'роман'	157,99
роман	133,89	näyttämö 'сцена'	156,97
поэт	130,29	runoilija 'поэт'	121,58
кресло	118,79	asunto 'квартира'	116,3
гость	111,34	miliisi 'милиционер'	109,26
дама	100,28	puhelin 'телефон'	109,2
луна	99,97	nikolajevits 'Николаевич'	98,41
голос	92,98	nainen 'женщина'	98,13
кабинет	87,46	näyttelijä 'актер'	89,67
николаевич	86,94	silmä 'глаз'	87,17
черт	85,77	huoneisto 'квартира'	86,84
дверь	76,84	työhuone 'кабинет'	75,79
трубка	72,52	nojatuoli 'кресло'	75,4

Таблица 5. Глаголы речи из русского и финского списков ключевых слов

Русский глагол	Частота	К	Финский глагол	Частота	К
ответить отозваться	404 103	282,13 164,76	vastata	541	315,75
заговорить сказать проговорить обратиться	137 756 29 42	121,01 15,49 32,30 24,56	virkaa sanoa	81 1176	70,43 69,72
вскричать воскликнуть прокричать крикнуть	76 108 37 79	116,09 69,79 49,28 13,14	huudahtaa huudahdella huutaa huudella voihkaista	218 16 220 25 14	121,72 13,46 16,36 17,96 6,92
спросить	333	64,97	kysyä	349	32,66
шепнуть шептать	47 49	40,54 22,85	kuiskata supattaa	104 27	44,27 27,71
добавить	62	40,08			
бормотать	40	27,23	mutista	65	31,54
попросить умолять	55 23	21,79 18,64	pyytää	165	43,62
продолжать	118	19,14			
усмехнуться	20	15,93			
			vannoa 'клясться'	39	34,53
возразить	19	14,47			
признаться	21	14,28	myöntää	31	18,65
			mainita 'упоминать'	44	8,20
			selittää 'объяснять'	69	9,09
			ulvoa 'выть'	25	10640
			änkyttää 'заикаться'	12	5,51

Таблица 6. Наречия в русском и финском списках ключевых слов

Русское слово	Частота	Финское соответствие	Частота
затем	129	sitten	748
тогда	89	heti	227
опять	163	taas	140
снова	5	taas	140
часто	5	usein	5

Таблица 7. Служебные слова в русском и финском списках ключевых слов

Русское слово	Частота	Финское слово	Частота
и	7640	ja	7363
что	2921	että	1509
впрочем	81	sentään	42
кроме	110	paitsi	87
для	126	vuoksi	39